PRIFYSGOL ABERYSTWYTH UNIVERSITY

Aberystwyth University

Privacy preservation in e-health cloud

Kanwal, Tehsin; Anjum, Adeel; Khan, Abid

Published in:
Cluster Computing

DOI:
10.1007/s10586-020-03106-1

Publication date:
2021

Citation for published version (APA):
Kanwal, T., Anjum, A., & Khan, A. (2021). Privacy preservation in e-health cloud: Taxonomy, privacy requirements, feasibility analysis, and opportunities. *Cluster Computing*, *24*(1), 293-317. https://doi.org/10.1007/s10586-020-03106-1

General rights
Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

Download date: 04. May. 2023

Privacy Preservation in E-Health Cloud: Taxonomy, Privacy Requirements, Feasibility Analysis, And Opportunities

Tehsin Kanwal[1], Adeel Anjum[1], Abid Khan[2]*

ABSTRACT

Electronic health records (EHRs) are increasingly employed to maintain, store and share varied types of patient data. The data can also be utilized for various research purposes, such as clinical trials or epidemic control strategies. With the increasing cost and scarcity of healthcare services, healthcare organizations feel at ease in outsourcing these services to cloud-based EHRs. That serves as pay-as-you-go (PAYG) "e-health cloud" models to aid the healthcare organizations handling with existing and imminent demands yet restricting their costs. Technologies can host some risks; hence the privacy of information in these systems is of utmost importance. Regardless of its increased effectiveness and growing eagerness in its adoption, not much care is being employed to the privacy issues that might arise. Privacy preservation need to be reviewed about the changing privacy rules and legislations regarding sensitive personal data. Our work aims at answering three major questions: firstly, how privacy models and privacy techniques correlate with each other, secondly, how we can fix the privacy-utility-trade off by using different combinations of privacy models and privacy techniques and lastly, what are the most relevant privacy techniques that can be adapted to achieve privacy of EHR on cloud.

Keywords: EHR; E-health cloud; Privacy; Generalization; Cryptography

1. Introduction

In E-health systems, Electronic Health Records (EHRs) is being increasingly espoused for collecting and storing various sorts of patient data containing sensitive information regarding patients' laboratory test results, demographics, personal statistics like age and weight and medications. EHR is a legal record i.e. its content and use are regulated. The content of the EHR cannot be changed any time (i.e. it is signed aftercare episode to prevent any change). EHR is also a logical record. The EHRs are operated by the EHR system, which is a collection of components that form the mechanism in which the EHR is generated, managed, saved, and recovered. This system involves individuals, health data, commands and techniques, applications, and communication facilities. The service providers or service provider organizations having the responsibility to manage the EHR are the data controller defining rules how others (data processors) process the information. The main objective of the EHR is to support patient's care and rehabilitation. There are also many secondary uses such as public health, clinical research, and statistics. We focus our work on the secondary use of EHR [1,2,3].

Abid Khan* abk15@aber.ac.uk [Corresponding author]

[1]Department of Computer Science, COMSATS University Islamabad, Park Road, Chak Shahzad, Islamabad 44000, Pakistan

[2]Department of Computer Science, Aberystwyth University, Aberystwyth SY23 3DB, UK

EHR system consists of various frameworks like information, research centre, radiology, and pharmacy, and so forth. The patients' information is stored at the public cloud is available and shared to different hospitals. Patients can give access to EHR records stored at the cloud. Health professionals can upload analysis reports, (for example, pathology reports) to the cloud, so it can be accessed to the specialists remotely for diagnosing the illness. Patients can manage their prescriptions and related data and provide this data to their healthcare providers. Health insurance companies can build the adequacy of their care management programs by offering some incentive added services and offering access to their group members[4].There are many implemented online EHR systems that range from initial and simple client-server based e.g. Vista, Open Vista, OpenEHR, to advanced functionality cloud-based EHR systems like CHISTAR, athenahealth, Inc. etc. [4-7]. The cloud-based EHRs approach is inspired by the principles of "*Availability*, *Scalability*, *Cost efficiency* and *Convenience"* to achieve the outsourcing paradigm. In cloud-based EHR, at one end, the dissemination of patient's data is greatly beneficial but on the other hand, it must be performed so immaculately that patient's privacy ought to be preserved. Privacy in cloud-based EHR is essential as many users in the public cloud are unknown and would be given access to EHRs for the quality of service to be provided to the clients [8].

Privacy is a complicated and multidimensional concept, defined in a legal, philosophical or indeed in a technical context. Information privacy is specifically related to addressing the problems regarding private information of a person and the exposure of this information, it can also be stated as "*the right of people, groups or organizations to decide for themselves when, how, and to what level information regarding them is transferred to others"* [20]. Privacy policies and various standards have been properly legalized in various countries to manage and secure the privacy of patients' records. "The *Health Insurance Portability and Accountability Act* (HIPAA)" the mostly adopted. The HIPAA states the US health-informatics' privacy rules. The EHRs storage is also following various standards, such as "*Health Level 7"* (HL7), to guarantee data privacy and security [12]. In recent years the EU data protection directive 95/46/EC, applied to EHRs data privacy is replaced by *"General Data Protection Regulation"* (GDRP). The goals of the GDPR are to secure consistent data protection rules in Europe, to propose reinforcement and modernize individuals according to their private data, and to improve the process of data flows [9,10]. Furthermore, outsourced data require extra vigilance other than following rules and regulations about data privacy and those specified in a model, named the CIA (confidentiality, integrity, and availability). It states the necessary directions for managing the security and privacy of data inside an organization [12,13]. We will also discuss the above-mentioned privacy-related rules and regulations in more detail in the privacy-preserving requirements section.

As it is difficult to maintain the balance between privacy and utility, protecting EHRs data is not simple. It is because when patients' data is ensured to be publicly available, it needs to be protected against several privacy threats e.g. patients identifying information disclosure, patients' sensitive information disclosure etc. At the same time, patients' specific information would be useful for subsequent analysis [14]. The adoptions of dynamic cloud computing environment to the privacy preserving techniques need some clarifications. (1) Security and

privacy issues that arise in cloud-based EHRs may not be the same as applied to privacy preserved EHRs using standard privacy techniques, it is due to the highly dynamic and distributed environment of cloud-based systems. (2) Cloud-based EHRs, highly vulnerable to privacy disclosures. For instance, a collision attack is possible on the data that are outsourced to the cloud and exposed to multiple users [15]. Due to the collusion of the cloud service provider and data users, a whole data set that is outsourced to the cloud and the privacy protection mechanism can be exposed [15]. Therefore, the privacy techniques that are secure in legacy systems for EHR systems might be susceptible to different kinds of attacks when introduced for EHR on the cloud. (3) Most of the privacy techniques and models that are applicable to EHRs in the cloud-based environment or standard computing environment for privacy models are almost same, though the settings change overall general mechanisms of underlying privacy techniques remain unchanged.

1.1 Motivation: Consider a scenario in which a healthcare organization outsource its data to the cloud for providing global information services and access at the merest cost. Hybrid cloud is an efficient framework introduced for providing secure data management. This cloud comprises of a private cloud, which is used to maintain sensitive information within the healthcare organization and a public cloud that stores the remaining part of the dataset. Specifically, the data publisher first divides the data into sensitive and insensitive parts and transmits them to private and public clouds. Next, a series of sanitization operations are performed on both parts to make the data secure. Then, authorized users may access the sanitized data e.g., medical practitioners or pharmaceutical companies for data analysis via query interface Next, a series of sanitization operations are performed on both parts to make the data secure [15].

The privacy concerns are frequently faced by the data owners and practitioners [17,18], despite the benefits of cloud-based EHR services. Privacy issues are approaching while outsourcing patient EHR to the cloud. It is because of the sensitive nature and "lawful and social" repercussions for EHRs data disclosure. A clear path is to encrypt the healthcare data before transmitting it to cloud [19, 20, 21]. Nonetheless, encrypted information processing isn't effective and is limited to specific tasks consequently making it unsuitable for EHR information with multipurpose uses [15].

For instance, homomorphic encryption can be a possible solution to this scenario. It is a powerful encryption technique that supports computations sans decrypting the input. While supposedly possible, applying homomorphic encryption may not be suitable in real life scenarios since it is computationally intensive [22]. Also, encryption introduces substantial overhead while answering the queries [15]. To overcome these limitations, privacy-aware anonymity-based approaches came into existence.

1.2 Our Contribution: The contributions can be summarized as below:

1. We review privacy-aware anonymity-based techniques for EHR.
2. We perform an in-depth comparative analysis of privacy techniques on basis of their merits and demerit.

3. We discuss privacy techniques, their utilization of privacy models and possible combinations of models and techniques. We also investigate different data types and their application in different privacy techniques.
4. We define a separate taxonomy for privacy-preserving requirements and provide a categorization depending upon their priority in e-health cloud.
5. We perform a comparative analysis of "privacy-aware anonymity techniques" based on the "privacy-preserving requirements" and highlight some e-health specific privacy techniques.
6. Finally, we explore the adoption of privacy-aware techniques for big data applications.

1.3 Organization of the paper: Section 2 presents a taxonomic overview of the related work in EHRs privacy at cloud. In section 3 a comprehensive comparative analysis of "privacy-aware anonymity-based techniques" has been performed. Privacy preserving techniques are evaluated based on their merits, demerits alongside their applicability to different data types. Also, highlights the utilization of privacy techniques and different privacy models. Section 4 provides a detailed taxonomy of privacy preserving requirements for EHR on cloud. In this section, privacy requirements are defined and discussed in detail. Section 5 presents a comparative analysis of "privacy-aware anonymity techniques" based on the "privacy preserving requirements". It will help us logically to find a better technique that is more appropriate when used for the outsourced data. Finally, Section 6 concludes the review alongside future directions.

2. Privacy Preservation in cloud-based EHRs

Our work is motivated by a few approaches that are utilized to settle security and protection issues of EHRs in the cloud (It will provide background knowledge about various existing approaches to protect the public, hybrid). We briefly review the relevant work on the privacy preservation of cloud-based EHRs in this section privacy of EHRs. Despite the existing solutions, privacy issues are major obstacles that are limiting the widespread adoption of public clouds across the globe. The main reason for this concern is that the information needs to be published to a broad and possibly anonymous set of receivers and sensitive data are hazardous for outsourcing to cloud so, there is an increasing need to investigate data anonymization techniques applied to this domain [8, 23]. A diverse set of techniques have introduced that claim to protect patients' privacy by applying various cryptographic and hybrid access control techniques [24-37, 54, 38-46] (we will briefly discuss these techniques in Section 2.1 and 2.2). However, most of these approaches have certain shortcomings (computational cost, require complicated key management systems and public key infrastructure (PKI)) whereby causing them less effective for the health data outsourced to the cloud [16, 55]. An immediate alternative to cryptographic approaches is a collection of Privacy-aware anonymity-based techniques e.g., partitioning based techniques [15,19] and differential [125] to the outsourced healthcare data [15]. Intel carried proof of concept to illustrate that the anonymization technique used in privacy models can be utilized in cloud computing to achieve anonymity [22]. A hybrid model of various anonymization techniques

like *k*-anonymity, *l*-diversity, and *t*-closeness is proposed to decrease the privacy disclosure risk. The privacy technique highly reduces the average re-identification risks [13].

In anonymity-based approaches, there exist different *privacy models, techniques, and algorithms* to preserve the privacy of microdata. Most of these privacy definitions were introduced for publishing patient records alongside satisfying some privacy and utility objectives. To comprehend the scenario, one must be able to understand the *relationship among the privacy models, privacy algorithms, and privacy-aware anonymity-based techniques.* Privacy models (e.g., *k*-anonymity, l-diversity, *t*-closeness etc.) are developed to work against privacy threats e.g. *identity disclosure, attribute disclosure or membership disclosure*. In simple words, privacy models draw a *baseline,* indicating the level to which they can assure the privacy of given individuals. *Privacy algorithms* follow certain privacy models and assure that the data can be modified in a way that may protect the privacy of concerned individuals [14]. Privacy algorithms make use of certain *Privacy-preservinge anonymity-based techniques* or (privacy techniques in short)[1] (e.g., g*eneralization, suppression, anatomy* etc.) to accomplish the most desirable trade-off between privacy preservation and utility.
Existing surveys [77,79,81,128,141,142,143,144] on anonymity-based approaches have performed just a comparative investigation of privacy models that are based on their properties and presented their internal implementation. These surveys require general guidelines about (i) how privacy models and privacy techniques correlate with each other and (ii) how can we improve the trade-off between privacy and utility by using different combinations of privacy models and privacy techniques (iii) what are the most relevant privacy techniques that can be adapted for cloud based EHRs.
In the literature, various approaches exist that claim to preserve the privacy of EHRs, when stored in the cloud. The taxonomy of privacy aware approaches in cloud-based EHR is given in Figure 1.

We categorize privacy techniques for cloud-based EHRs into *cryptographic*, *cryptographic policy aware hybrid* and p*rivacy-aware anonymity-based techniques*. Cryptographic and policy aware hybrid techniques are briefly covered in Section 2.1 and 2.2. while Section 2.3 provides an introductory note on privacy aware anonymity-based techniques. We provide a comprehensive review of these techniques in Section 3. Cloud deployment models in EHRs context are given in Section 2.4.

2.1. EHRs Privacy: Cryptographic techniques

We will present the related work on cryptographic techniques in this subsection. However, we suggest interested readers to [25] for an in-depth analysis of these techniques.

[1]We will use Privacy aware anonymity-based techniques, privacy techniques and anonymization techniques interchangeably throughout the article.

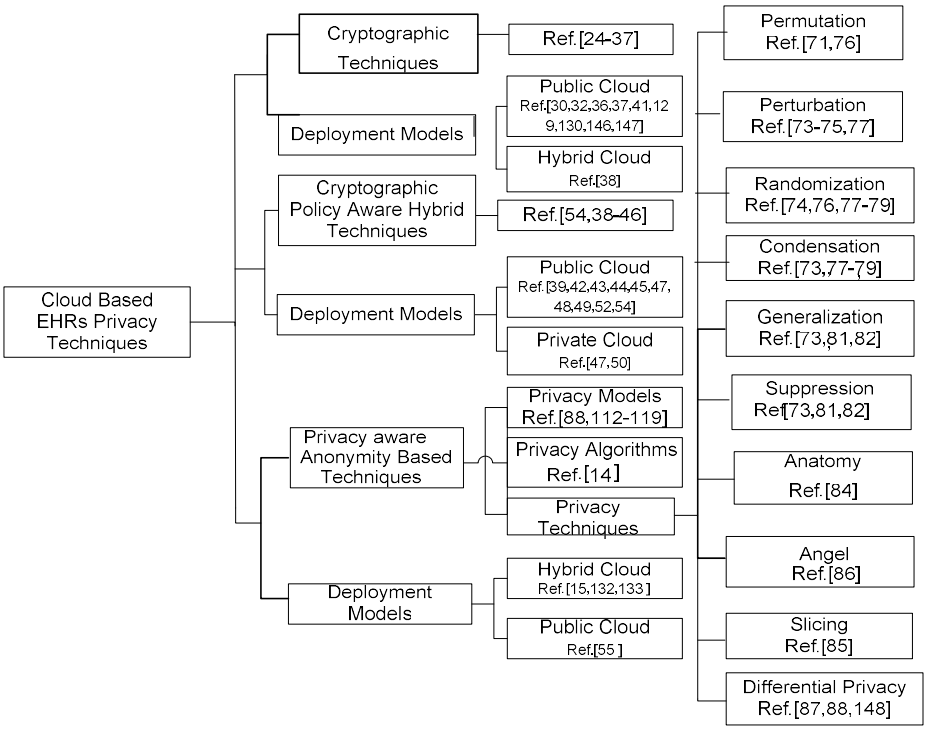

Figure 1: Taxonomy of privacy aware approaches in cloud-based EHR

recommended by NIST). **Public key encryption (PKE)** employs both private and public keys. The encryption through PKE is safe. However, it is computationally not suitable. It uses the combination with the SKE, RSA and elliptic curve cryptography (ECC) techniques [25].

SKE-Based approach to secure EHR in the hybrid cloud: Patients' data privacy is claimed to be achieved by encrypted EHR, using the SKE hybrid cloud. Patients' data is stored in the hospital's private cloud and the public cloud for health care providers. Health data can only be decrypted using the private key. It split randomly, and one part is stored at the hospital server and another part is at the patients' smart card. Although they achieved patients' data ownership, however, the patient's privacy remains to be addressed [38].

Attribute-based encryption (ABE), In ABE encryption, and decryption is carried on based on the attributes of users. For decryption user attributes and decryption keys are necessary. ABE allows users to selectively distribute encrypted data and provides fine-grained access [26, 27]. An ABE-based access control mechanism is used to provide the anonymity of the users by saving the PHRs at the cloud. The proposed approach is effective against the replay attacks and shares the keys in a decentralized manner [147]. Ciphertext-Policy Attribute-Based Encryption (CP-ABE), this encryption technique is performed on access policies. CP-ABE practice is bounded for the reason to specify access control policies. Management of user's attributes is also a problem in CP-ABE [28]. Key policy attribute-based encryption (KP-ABE), the access policy is linked with the private key and encrypted text is marked with a set

of identifying attributes. Decryption is only possible if access policy and data attribute matches. Various trusted authorities create the users key in multi-authority attribute-based encryption (MA-ABE). These authorities are also responsible for governing users' attributes and the user can get only a part of the secret key from trusted authorities [29]. It is applied to the secure and scalable system to share EHR in the cloud in [30]. In Searchable Encryption, the search operation is performed over encrypted data without disclosing any information about the contents. It was practically applied using the symmetric key cryptography [31]. It is applied to search for EHRs in [32]. The identity-based encryption (IBE) may use any string e.g. name or email address as a public key. A trusted party issues decryption key in IBE [33]. hierarchal IBE (HIBE), the complex task of private key generation is handled within the hierarchical form. The HIBE approach is used for the protection of EHRs in [34]. Fully homomorphic encryption (FHE), permits computation (random number of additions and multiplications) over encrypted data without decrypting it [35]. Another variation of it is "somewhat homomorphic encryption (SwHE)" that limits the number of encryption operations with the help of evaluating circuits of a specified depth. SwHE is applied to the cloud to perform computations over encrypted patients' data in [36]. In another work, the cloud-based privacy-preserving system is described. Privacy of patients and mobile health service providers are protected with the help of identity-based encryption (IBE). Homomorphic encryption is used to protect data transmission from healthcare providers to the cloud in [37].

2.2. EHRs privacy: Policy aware hybrid techniques.

We present a concise analysis of hybrid approaches in this section. Hybrid schemes are a combination of cryptographic and access control based techniques to preserve the privacy of EHR at a cloud.

SKE-Based scheme for EMR sharing in cloud: In this approach, to achieve unlikability between patients and EMR in a cloud environment, patients EMR are encrypted through SKE and saved anonymously. Digital signatures are used to protect Personal Health Record (PHR) and stored in the cloud. A smart card that contains identity seed (SID) in PHR is used to access the patients' health data. Personal identity is stored in two parts independently to prevent illegal patients' data access [39]. SKE based approach is also used in [38] to achieve forward data secrecy; data unlikability and integrity for PHR stored in the cloud.

Patient-centric role-based EHR in the untrusted cloud is a reference model for preserving privacy in healthcare applications, by assuming an untrusted cloud. In this model, a patient-centric and RBAC model is considered to achieve anonymity. The group signatures scheme is used for the authentication and integrity of the EHRs data. In this scheme, a group member anonymously signs a message on account of all the group members. There is not any specification of privacy approach is given to achieve anonymity [41].

Attribute-based privacy preserving EHR system: In this approach, Secure Channel Free PKE with Keyword Search (PEKS) is introduced that enables the users to search for a matching keyword without exposing the original keywords to the server. The system provides

confidentiality of health data with the help of encryption and claims to provide privacy to the extent that the cloud server will not be able to learn contents from ciphertext and keyword searches [42].

EHRs system using (CP- ABE) in cloud: In this approach, healthcare providers use the CP-ABE scheme using public keys and private keys for encryption and decryption. Healthcare providers share one public key for encryption, so that there may be no need for PKI for distribution and management of keys. The secret key for health care providers can only decrypt the specific ciphertext. If a secret key is exposed then EHRs data, can be decrypted with that specific key will be compromised and other EHRs remain safe [43].

CP-ABE based access control approach: The scheme permits users to access health data based on their access rights using CP-ABE. Secure communication has performed between Patient and e-health CSP with the Identity-Based Encryption (IBE) [44].

Data sharing in cloud with privacy-preserving access control: The proposed system separates data protection from CSPs and unauthorized users. Commutative encryption with multiple layers, protect data against CSPs. In this approach, an access control mechanism protects information from malicious users. It uses SKE based encrypted data is stored in the cloud. The symmetric key is re-encrypted using commutative multi-layered encryption. Access control policies for different granularity levels are by the data publishers. [45].

Situation-based access control: It is a concept-based model that describes a patient's information access scenarios through situation schema using specific patterns and relations. Sit-BAC scenarios provide a basis for the preservation of the patients' privacy. Additionally, it represents a situation where there is a conflict between the data requestor or role and the requested data in the same organization. Its focus is on medical domain [46].

Role and time-based access control: The method is useful while saving the encrypted EHRs at the untrusted cloud service provider.it resolves the problems of key distribution amongst legitimate users. SKE based approach is used for protection of EHRs. [146].

Pseudo anonymity policy based authorization: The linkage of different datasets of a single patient is created through the Pseudo Anonymity Policy-Based Authorization approach. Moreover, by tracing the flow of health data. It allows the patients to audit access activities for EHRs. Anonymity is provided with the help of pseudo-anonymity service. Health data is also encrypted with a different key before it can be stored at the third-party server [47].

Pseudo anonymization for secondary use of EHRs. A scheme is given to protect the patients' privacy in ordinary healthcare activities and secondary usage of EHR for medical research. It uses the Certificate Authority to issue certificates to doctors, pharmacies, Researchers, and Insurance companies. Patients generate their Master Secret Key (MSK). All private information in EHR and communication is preceded by it. Pseudonym scheme using MSK replaces patient's real ID by pseudonym PID (each time when the patient visits a doctor), other patients' sensitive information is also encrypted by MSK, when communicated with other entities at cloud [48].

SAPPHIRE: It introduces the secret usage of cloud services into existing cryptographic based mechanisms. To enable research and storage of public health data, SAPPHIRE preserves the privacy of the anonymous submitter. It can be achieved using an anonymous identifier or pseudonym that is used for de-identifying the user's Personal Identifying Information (PII). Pseudonym ID is used to submit, or access data collected at the cloud. [49].

CAM-Access of EHRs with privacy and auditability Solutions claim to provide protection in the mobile healthcare system at private cloud. It gives protection for information storage and retrieval, particularly at crises. Unlinkability is provided by the integration of pseudorandom number generator with the key management process. A secure indexing method for privacy-preserving search and access patterns based on redundancy is also introduced with the help of ABE, RBAC, and Audit ability. They try to control the odd behavior of any user, in both normal and emergency situations [50]. Likewise, Pseudo anonymization is also claimed to preserve the privacy of EHR in works given in [51, 52,53].

Hybrid approach for medical data sharing: In [54], an effort is made to provide privacy of patient in shared medical data in case when there is the possibility of conventional policy-based models privacy breach. A hybrid solution of statistical analysis and cryptography is provided to ensure the maximum data usage with privacy preservation. Its main components are: (1). Vertical data partition that is performed at owner side provides fixed vertical partition of medical data. (2). Data merging process for health dataset access at the authorized recipient side. (3). Checksum based integrity checking and probability mechanism is performed at both the local and global levels (4) A hybrid search using keyword search across plaintext and cipher text is used.

Privacy aware content disclosure for EHRs: To preserve the privacy of cloud-based EHR, a hybrid approach is used that combines CP-ABE and *k*-Anonymity for anonymization. The authors in [55] have claimed to achieve fine-grained access control.

A precise overview of privacy preserving techniques for EHRs on cloud is given below in Table 1.

Table 1: Privacy Preserving Techniques for EHRs on Cloud

Privacy Techniques	Findings/Remarks	Cloud Model
Multi-Authority Attribute-Based Encryption (MA-ABE) [30]	*The scheme is complicated and computationally expensive. It does not provide privacy and requires a fine-grained access control mechanism for the public cloud.*	Public
Searchable Encryption and Symmetric Key Encryption (SKE) [32].	*Privacy-aware of access control for EHRs needs flexibility. The scheme works fine for EHRs search ability however not for refined access control.*	Public
Symmetric Key Encryption (SKE) and Digital signatures [39].	*In this approach, proper access control has not implemented as given after a simple login password in access rights. To provide anonymity and unlikability, some more useful privacy alternatives can be used with access control in the hybrid cloud*	Public

Symmetric Key Encryption (SKE)[40].	Need to incorporate utility aware mechanisms and access control list mechanism is not flexible in hybrid cloud need improved audit. Instead of SKE anonymity-based mechanism can be used to preserve privacy and improve utility of patient data	Hybrid
Attribute-based encryption and public-key encryption [42]	*Attribute-based encryption is expensive in terms of high runtime when the number of attributes in health data increases. Attribute-based encryption (ABE) can be used with a less expensive policy anonymization solution to preserve policy privacy from CSP. Furthermore, key management and distribution need PKI.*	Public
(CP-ABE) + Identity-Based Encryption (IBE) [44].	*The solution requires improvement in case of privacy of high dimensional health data. Also, CP-ABE will not remain efficient for health data with a large number of patient's attributes.*	Public
Symmetric Key Encryption (SKE) and Commutative Encryption-Based Access Control [45].	*The proposed approach does not support cloud scalable infrastructure for access control policy enforcement. It needs protection against collusion between CSP and malicious users. Overall the solution is not complicated, however, access policy creation and enforcement mechanisms can be further improved at the public cloud.*	Public
Situation-Based Access Control (Sit-BAC)	*The proposed solution provides a structured specification of patient data access scenarios with various situation models. However, it is not a scalable and fine-grained access approach.*	Public
Pseudo Anonymity and Certificate Authority Based approach [48].	*The solution is expensive because of the use of the Certificate Authority for the patient's privacy. EHR data linkage privacy attacks are possible. Moreover, the fine-grained access control mechanism for both primary and secondary use of EHRs data can be used in a hybrid cloud*	Public
Hybrid Pseudo-anonymity, ABE, and RBAC[50].	*The solution is complicated to provide privacy in the mobile health care system. There can be flexible access control with enhanced usage.it can also be improved with hybrid cloud and use of the anonymity-based solution for storage and retrieval.*	Private
Statistical analysis and Cryptography Based approach [54].	*Use of both SKE and PKE cryptographic techniques make solution expensive in EMR data context. Vertical data partition can be replaced with privacy techniques that automatically perform all partitioning and merging of the EMR data set.*	Public
CP-ABE and *k*-Anonymity [55].	*CP-ABE is not suitable for health data with a large number of attributes. Policies, in the plain form, are vulnerable in the public cloud. This solution is not scalable in data attributes. The anonymization method can also be further improved.*	Public
RBAC, AES-256, SSO and MAC [58]	*The proposed approach provides interoperability and scalability with reduced costs. However, interoperable and fine-grained access control mechanism has not given.*	Public
Personalized and differential privacy-based approach [15]	*In this approach, the absence of proper access control provides a source of internal and external collusion between users and CSP even on sanitized data in the public cloud. It can be improved with the inclusion of privacy technique that enhances data utility.*	Hybrid

Map Reduced data partitioning approach [132]	*It is not a privacy-aware solution as Its focus is on the automatic division of computation tasks by data partition technique. Map Reduced technique can be used with efficient privacy-preserving mechanisms as an enhancement mechanism*	Hybrid
Map Reduced data partitioning approach [133]	*It is a restricted solution for multipurpose usage of stored data at the public cloud. In this approach, there is a need for an efficient privacy-preserving mechanism.*	Hybrid

2.3. EHRs Privacy: Privacy-Aware Anonymity-Based Approaches

The taxonomy in Figure 1 divides privacy aware anonymity approaches into; Privacy Models, Privacy Algorithms, and Privacy Techniques. Privacy models are basically developed to provide defense against privacy breaches or threats. Privacy threats as given in [14,128], are basically divided into three types:

- **Identity disclosure:** if an attacker associates an individual identity with his/her record in a published dataset, it causes an identity disclosure.
- **Attribute disclosure:** In this type of privacy disclosure**,** an individual sensitive attribute is disclosed by an attacker with the help of externally available information.
- **Membership disclosure:** It occurs when an attacker can accurately guess with a high level of probability that an individual's record is present in the published data.

Privacy Algorithms use certain privacy models and privacy techniques e.g., *generalization, suppression, anatomy* etc. Privacy techniques have various anonymization mechanisms to transform data into an anonymized form. Privacy-preserving techniques, such as permutation, perturbation, generalization, suppression, anatomy is used for sensitive data protection when it is published. A brief overview of privacy threats, models, algorithms and privacy techniques given in Table 2. However, a more elaborated details of privacy threats in terms of privacy models and techniques will be given in Table 5 in next sections.
We present a detailed description and discussion about "Privacy preserving techniques" in the next section.

Table 2 Privacy threats, Privacy Models, Algorithms and Techniques Summary

Privacy Threats	Privacy Models	Privacy Algorithms	Privacy Techniques
Identity disclosure	K-Anonymity [112], k-Map [56] (1, k)-Anonymity [57] (k, k)-Anonymity [57]	k-Minimal generalization [62] Incognito [63]	Generalization and Suppression.
		Mondrian [63,64], Greedy [65] TDS [66], KACA [67]	Generalization.
Attribute disclosure	l-diversity [113] (a, k)-Anonymity [68] t-Closeness p-sensitive-k-anonymity [59]	Incognito(l-diversity) [113], Incognito(t-closeness) [114], Incognito (a, k-anonymity) [68]	Generalization and Suppression'.
		Mondrian (l-diversity) [69] Mondrian(t-closeness) [84]	Generalization.
Membership disclosure	d-Presence [117] c-confident d-presence [60]	SPALM [117], MPALM [117], SFALM [70]	Generalization.

2.4. Cloud Deployment Models: EHRs Context

Privacy-preserving techniques for cloud-based EHRs based on the cloud deployment model will be analysed in this subsection. We have observed that most of the privacy preservation work is performed at the public cloud, whereas the hybrid cloud has been limitedly used for EHRs data privacy protection. Cloud deployment models in EHRs privacy context are detailed below.

2.4.1. Public cloud deployment model

The public cloud infrastructure is accessible to public users and is controlled by a cloud service provider. The electronic health records can be shared with various participating entities, for example hospitals, clinics, pharmacies, insurance companies, and clinical laboratories in a public e-health cloud. In public cloud, the EHRs are kept at the CSPs' managed off-premises servers [12].

While residing in a public cloud, the EHRs' security and privacy are at high risk of malicious attacks [135], from external and as well as internal entities.

Consequently, mechanisms are needed to achieve privacy and to ensure confidentiality of EHRs. Consequently, mechanisms are needed to achieve privacy and to secure the confidentiality of EHRs data. Strong encryption techniques and effective signature verification schemes are previously employed but trivial work is performed for the protection of EHRs privacy through anonymization-based techniques. Majority of the security and privacy protection work is done at public cloud for cryptography techniques [30,32,36,37,41,129,130,146,147] and cryptographic policy aware hybrid techniques [39,42,43,44,45,48,49,52,54].

2.4.2. Private cloud deployment model

The private cloud's infrastructure is supposed to be administered by the healthcare organization only. The healthcare organization or a trusted third party may partake in private cloud's management either on or off premise. The cloud's infrastructure units; storage and processing units are generally managed by some designated third party of the hospitals [12]. The electronic health records stored at private cloud are meant to be more secure than that at other deployment models, in the absence of public internet. Only trusted healthcare parties can access the EHRs in the private clouds. In [13, 61], the researchers have proposed some cryptographic hybrid techniques in this regard so far.

2.4.3. Hybrid cloud deployment model

The hybrid cloud is a unification of the public and private clouds. In e-health cloud, hybrid cloud deployment model is most significant. Healthcare organizations with insufficient physical resources can store EHRs and other medical data. To use the maximum benefit of cloud computing and to overwhelm the shortcomings of private and public cloud, hybrid cloud assures an effective and robust resolution for prospective healthcare applications [12,139,140]. Privacy Preservation of EHRs is a major issue in hybrid clouds and requires innovative solutions for solving the privacy challenges in the e-health cloud. Privacy-

Preserving Anonymity Based Techniques are applied to the hybrid cloud [15,132,133] and public cloud [55].

3. A comprehensive overview: privacy preserving techniques

This section will provide a review of various "privacy preserving techniques", their merits, demerits and data applicability. The section ends with a brief discussion about privacy models and privacy preserving techniques by evaluating their different combinations to find a better combination to achieve the best possible level of balance between privacy and utility.

In Table 3, we have summarized "privacy-aware anonymity-based techniques" based on their merits, demerits, and applicability to various data types. In Section 3.1, we briefly overview each technique while in Section 3.2, we provide a detailed account of the application of each data type to a given privacy technique.

Table 3: Comparative analysis of Privacy-Aware Anonymity- Based Techniques

Privacy Techniques	Merits	Demerits	Data Applications
Permutation [71,76]	• It is useful to be applied in theoretical data mining experiments. • It can be applied in combination with other data anonymization techniques e.g. slicing.	• It produces inaccurate data values so causes data utilization loss. • Not suitable for real word -data anonymization. • In the case of SA of medical data (e.g. Disease), it is not relevant a technique as it is only applied to numerical values of SA.	Relational Data [71.76,85,90] Transaction Data (set-valued data) [89] Graph Data [91]
Perturbation [73-76,80,120] **Adding Noise** **Data Swapping** **Synthetic data generation**	• Data perturbation is simple and effective. • Perturbation preserves statistical information. • Mostly used for statistical disclosure control.	• Perturbation provides low level data privacy and at specific conditions it is not difficult to breach the privacy protection. • In perturbation data dimensions are independently treated so lost the correlation. • These techniques produce synthetic data so it's of no use to data recipients. • Algorithm level implementation is hard as it must design new distribution specific algorithm. • Privacy threats that are supposed to handle are unlikely to occur.	Relational Data [73- 76,80, 120] Textual Data [92]
Randomization [74,85]	• Randomization preserves some statistical properties such as mean and correlation.	• Randomize results are approximate results. • There is huge information loss and so cause low data utility.	Relational Data [74,78] Textual Data [92]
Condensation [73,80]	• Condensation works on original data set instead of data distribution. • Condensation does not require a redesign of new problem specific Algorithm for data mining.	• Condensation produces synthetic data so for real world data recipient it's not useful. • Condensations lose data statistics when working in the conversion of large groups of data into the small condensed group.	Relational Data [73,80] Multi Graphs [93]

Suppression [73,75,80]	• Privacy preserving by hiding data with'' *''' • Easy to implement • The inferential attack is impossible. • Statistical characteristics discovery is impossible. • Provide optimal solution for privacy. • Unauthorized access to the data does not cause any privacy breach.	• Hard to decide which values will be suppressed and either partial or complete suppression. • Extra storage space utilization. • No protection against background Knowledge. • It does not work well on high dimensional data. • Excessive use reduces data utility.	Relational Data [73,81,82] Transactional Data (set-valued data) [99,102] Text Data [100] Trajectory Data [101]
Anatomy [74,83,84]	• Improves data utility and both SAs and QIs are in their original forms. • Outperforms in terms of no information among generalization, suppression and slicing • Aggregate queries requesting domain values of SA and QID can be answered accurately in anatomized tables.	• In Many data sets, it's hard to separate QIs and SAs. • As QI and SA are in unmodified form adversary can correctly identify the presence of individual in microdata and leads to membership and may be identity disclosure • Attribute correlation is lost by QIs and SAs separation. • Application of techniques such as classification, clustering to published data is not clear as data is published in two table formats. • Background knowledge attack is one of the great challenges for anatomy.	Relational Data [74,83,84]
Slicing [85,128]	• Provides better utility of the data. • Provide protection against identity disclosure. • Provide Highly correlated data. • Better performance • High dimensional data anonymization can be performed. • It can be applied to the dataset where QIs and SAs are not clearly separated.	• Correlation rules and its implementation is challenging. • Vertical and horizontal partitioning requires extensive resource utilization and storage • Provide protection against membership and attribute disclosure but about identity disclosure, it is not clear.	Relational Data [85] Transactional Data (set-valued data) [127]
Angel. [86,128]	• Applicable to anonymization technique e.g., k-anonymity, l-diversity, t-closeness, etc.). • Preserve significantly more information than traditional generalisation. • It easily advocates the marginal publications.	• The existence of large number of sensitive values in QI-groups causes problem especially when sensitive values are quadratic to the number of records in QI-group. • Aggregate queries results in higher average error for many sensitive values in qi-group.	Relational Data [86] Transactional Data (set-valued data) [127] Marginal Publication [86]
Differential privacy [87, 88,148]	• It can answer aggregate queries by preserving privacy; the objective is to minimize absolute error to achieve differential privacy. • It can be applied to both interactive and non-interactive methods.	• The utility of data can be affected by a large amount of noise added to the database. • There are no practical guidelines to measure the value of epsilon.	Relational Data [87,88,107,148] Transactional Data (set-valued data) [103,108] Graph Data [105,106,149-151] Big Data [21,104.125]

3.1 Privacy preserving techniques

Privacy Preserving Techniques are used for protection of patients' sensitive data when it is publicly published. Some techniques, such as permutation, perturbation, condensation adopt data mining strategies. However, some are purely developed to handle microdata privacy for the publication of anonymizing data in an unperturbed form (Generalization, suppression, anatomy etc.). In the following section, our focus is to comprehend each technique with a goal to demonstrate how each technique operates. Moreover, a comparative analysis is also performed based on the merits and demerits of anonymization techniques.

3.1.1 Permutation: It means rearranging the data values after partitioning into groups of values e.g. sensitive values in microdata. The permutation is basically a method for handling numerical Sensitive Attribute (SA) so in the case of categorical SA, permutation does not seem a relevant technique. The permutation is considered interesting for theoretical data mining experimentations after anonymization. Permutation has the disadvantage of producing fake values. Inaccurate values cause data utility loss. Permutation is not considered to be a suitable approach for sanitization of real-world data. Furthermore, Permutation alone cannot be used for practical purposes after data publication [71, 72].

3.1.2 Perturbation: Perturbation techniques are divided into two basic categories: *Input perturbation techniques*, in which data are randomly modified and answers to the queries are computed using the modified data. In *Output perturbation,* correct answers to the queries are computed exactly from the real data but some noise is added, and that version of results is reported. For private data publications, input *random perturbation* is used. It is achieved by random replacement of a sensitive attribute value by another sensitive attribute value in its domain. With a given retention probability *p*, if the coin heads then original value is retained, otherwise, the value is replaced by a random value in its domain. Three methods are used for data perturbation

(a) Additive noise: In this method, some noise is added or multiplied to each numerical SA values in microdata table. Perturbation preserves some statistical properties e.g. mean and correlation, but it also generates some fake values. Additive noise may become vulnerable and can cause privacy breach. Multiplicative noises do not have such drawbacks.

(b) Data swapping: These methods exchange the SA values among the records. Values to be exchanged should belong to the same attribute domain, though it does not change the domain of attribute values, but the combination is changed. Data swapping techniques can be applied to for anonymization of numerical attributes and categorical sensitive attributes.

(c) Synthetic data generation: In this method, a mathematical model is constructed that uses the original data to generate synthetic records. As, the released data preserve the unique features, although, they do not reveal real data. The integrity of data at the record level in these methods is not supported. [73-77,120].

3.1.3 Randomization: Randomization is a variation of classical perturbation technique. *Randomization* modifies the data values so that the data produced are a distorted version of original data. In response of randomization, the data are swapped so that it cannot be judged whether the data contain true or false information with possibilities better than a pre-defined threshold.This technique can be used at statistical attributes. In this approach, data are

perturbed with an appropriate level of noise, randomly chosen from a distribution, which can either be added to or multiplied by the original value of each attribute.

Randomization approach is simple and there is no need of a trusted server to hold original microdata, instead a perturbed copy of original data is released. The advantage of this technique is that it preserves some statistical properties such as mean and correlation, but it also generates some meaningless values, so the results are approximate and have huge information loss. Randomization deteriorates the use of original data due to the addition of unnecessary noise.

Randomization can better preserve utility in several different aspects of distribution reconstruction, accurate results of aggregate query answering, and correlation among attributes for the same privacy requirements. Utility loss is noticeably smaller than that of generalization or anatomy-based approaches. With the increase in the amount of available data, the effectiveness of randomization increases which is not the case in generalization and anatomy [74, 76, 77, 78, 79].

3.1.4 Condensation

In condensation approach, records are condensed into multiple groups or clusters and from each group, some statistical information is extracted. It is a statistical approach in which synthetic data set is generated based on group statistics. Synthetic data have same aggregate distribution as of original data group. Each cluster has some fixed size based on privacy requirements or level of underlying privacy principle. The amount of privacy achieved depends upon the level of underlying privacy principle. The more substantial amount of data is dropped because of the condensation of a greater fraction of records into a single demographic group. [73, 70,80].

3.1.5 Generalization: In generalization, the specific values of QI attributes (Quasi-Identifiers) are replaced with a more generalized range of values. The objective is to increase the uncertainty. Therefore, it becomes difficult for adversaries to link an individual to a record or his/her sensitive information. Generalization can be applied on both categorical and numerical attributes. Generalization continues until the microdata table satisfies some anonymization properties according to underlying privacy model e.g., k-anonymity, *l*-diversity, *t*-closeness. There are four types of generalization.

(a) Full-domain generalization, In the generalization method, the value of QI attributes must be generalized such that the whole domain values are generalized. It operates for the same range of numerical values**.** In the case of the categorical attribute, generalization is employed at the same level of the taxonomy.

(b) Subtree generalization, where a child node in the taxonomy tree structure is generalized to its parent node, all child nodes must be generalized to that parent node.

(c)Cell generalization, where a single cell of a record is generalized as compared to full-domain generalization where all tuples in that domain in the dataset are generalized.

(d) Multidimensional generalization, where the record is generalized by the combination of quasi-identifiers with different generalized values, creating a different generalization for different combinations of values of QI attributes [73, 74, 81, 82].

3.1.6 Suppression: In suppression, the value of the attribute is changed by the special character * so that it will not reveal any information. Suppression can also be linked to

generalization as a special case. There are different types of suppression exists in literature details are given in [73, 75, 82].

3.1.6 Anatomization: This approach has reduced the shortcomings of generalization and improved the level of utility in data publishing. Anatomy split the microdata table into two tables: A Quasi-Identifier Table (QIT) and a Sensitive-Identifier Table (SIT), which separates quasi attributes from sensitive attributes values. It does not transform the quasi attributes and the sensitive attribute values but disassociates the relationship between the two by separately releasing QIT and SIT. The QIT and SIT both are linked through Group ID. In this approach, each group has the same value of Group ID in both QIT and SAT tables. It connects the values of the sensitive attributes in the group. Each group has different sensitive values and each distinct value occurs precisely once in the group. When we apply the generalization approach, attribute domain values are missed, without supplementary knowledge, the uniform distribution theory is used to answer a query about domain values [74, 83, 84]. Consequently, the anatomy is viewed as a better approach than generalization.

3.1.7 Slicing: In slicing, the first step is the partitioning of attributes into columns. Each column may contain single attribute or subset of attributes; this is called *vertical partitioning*. Slicing also partitions records into buckets. This is called *horizontal partitioning*. attribute values are randomly permuted within columns to break the linkage among different columns. [84, 85].

Slicing is a novel technique proposed to provide a better solution for the problems which are not covered by traditional technique. It reduces the high dimensionality of data by splitting different columns to break association among them. Better data utility is produced by slicing as it groups highly linked attributes together. Slicing protects privacy by disrupting the association between uncorrelated identifying attributes.

3.1.8 ANGEL: It is a new privacy technique, designed to enhance the flaws in prior techniques, particularly for generalization. Angel improves considerable information loss by preserving more information and at the same time, it maintains the same privacy level. Angel solves the complex problem of marginal publications too.

Angel first divides the table into batches (simply make a group of records obeying *l*-diversity for sensitive attribute e.g. disease) in patient's microdata table, called batch table. The batch table (BT) gives the summary of the disease statistics of each batch. Next, the Angel creates another partitioning of the table in form of buckets (groups of tuples that need not be *l*-diverse). Moreover, the Angel generalizes the tuples of each bucket into the same form, to produce a generalized table (GT). The Tables, BT and GT, are the final relations released by ANGEL [86].

3.1.9 Differential Privacy Differential privacy is an advanced privacy technique for Statistical Information Disclosure (SID). It attempts to ensure that individuals expect minor disclosure risk of their sensitive attributes when they agree to be the part of a database. Differential privacy achieves it with the requirement that the addition or elimination of a individual's record does not modify the output of the function.

As for Differential privacy, by definition, does not distinguish between interactive and non-interactive mechanisms so, it can be applied to both the sanitized publication and the perturbed queries scenarios [87]. Differential privacy enables the statistical analysis of sensitive information and providing strong privacy guarantees yet in the presence of random

auxiliary information. The Laplacian noise that is added is random and can prevent adversarial linking attacks effectively while at the same time preserves data utility. It is claimed that it does not lose the integrity of the data as compared to its counterparts [87, 88, 148].

3.2 Privacy-aware anonymization techniques and various data types

The data anonymization techniques proposed in the literature can be investigated in several dimensions and data type is one of them. In this subsection, we explore the applicability of each anonymization technique to a variety of data types. When we talk about the data anonymization techniques and their applicability to different data types, we have an ample amount of literature in this area. Also, it covers a wide range of domains including data mining, healthcare, and social networks etc. We present some of the work that from our perspective, will be sufficient for the reader to have an idea of data type applicability to different privacy preserving techniques.

Before delving into the details, we explain several data types with the help of taxonomy as shown in Figure 2. We can categorize the data types into (a) Structured data, which represents relational data, traditional text and numerical information (b) Unstructured data, that includes variety of data e.g. transactional data, trajectory data, images, audio and video files, textual data(e.g. PDF, word documents, presentations, etc.), emails and human language (c) Semi-structured data, such as XML, HTML, RSS feeds and multigraphs (d) Big data, it comes from heterogeneous sources and generally is in three types: structured, semi-structured, unstructured and streaming data [109-111].

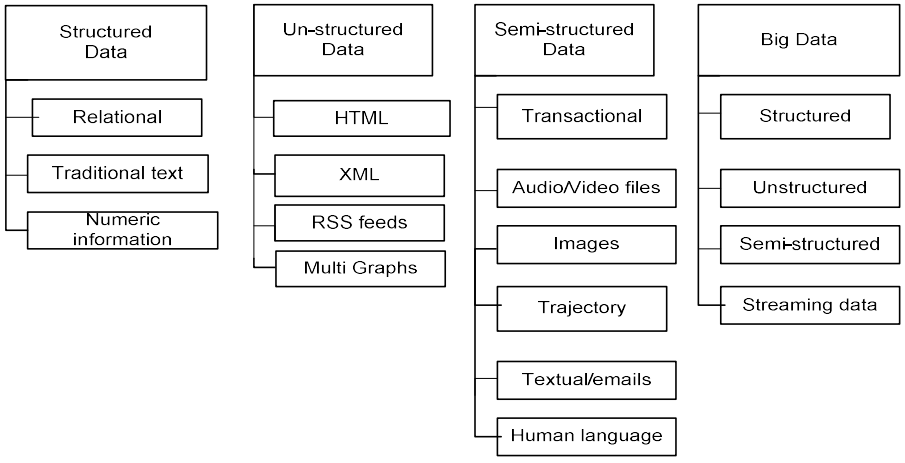

Figure 2: Taxonomy of various data types

Please note that we have intentionally categorized big data separately because the privacy techniques applied to big data differ from the traditional ones.

In Table 2, we have presented privacy aware anonymity techniques with their merits and demerits. Privacy techniques in terms of data type applicability are given in the last column of Table 2. Since the merits and demerits of privacy techniques are self-explanatory, a review is performed on the applicability of privacy preserving techniques on different data types. A permutation is used in most of the data types including structured, unstructured, and semi-structured data. In [89], it is applied in combination with generalization to anonymize sparse high dimensional data (transactional data). In the case of structured data, it is used in relational microdata including numerical and categorical attributes [90, 85, 71, 76]. It is also applied to semi-structured graph data in [91]. Perturbation is successfully applied on structured data in [73-76, 80,120]. For unstructured data, the perturbation is employed in [92] to anonymize text data. It is used in conjunction with generalization to hide the personal identity information in the textual data. Randomization is used in structured and unstructured data in [92, 74,78]. Privacy-preserving data mining is performed by utilizing a combination of randomization and suppression-based algorithms. It is applied on text and binary data by using a sketch-based approach. It is claimed that the proposed technique is extremely effective for high-dimensional sparse datasets [92]. Condensation is applied to structured data in [73, 80]. It is also applied to semi-structured data in [100] that deals with the anonymization of multigraphs. Here, condensation is used in combination with the k-anonymity based algorithm. Authors of [93] claim that this is the first work which examines the anonymity problem in the context of multigraphs. Generalization is widely applied to almost all data types except semi-structured data. Generalization is applied to structured data in [74, 94, 81, 82]. In [94], it is successfully applied to anonymize both relational and transactional data. Privacy of set-valued data also called transactional data, is preserved in [95, 96] by applying full subtree generalization. In [97], an approach for anonymization of text data is proposed which uses multidimensional generalization to enforce *t*-plausibility. A hybrid scheme that couples top-down specialization (TDS) and bottom-up generalization (BUG) is applied to Big data at the cloud using map-reduce for scalability in [98]. In [15], Generalization is used in conjunction with differential privacy and personalized privacy [119] for achieving the privacy in big data on the cloud. Suppression is used in structured data in [73,81,82]. In the case of unstructured data (transactional data), a combination of suppression with full subtree generalization is used to achieve *km*-anonymity. In another work for transactional data privacy, a novel privacy notation (*h, k, p*)-coherence use suppression to achieve coherence [99,102]. It is used for anonymization of items from the text document in [100]. Privacy preservation for trajectory data is also studied in [101] that use local suppression.

Anatomy is applied to only structured data in [74, 83, 84] for preserving the privacy of individuals in microdata i.e., relational data about individuals. Use of slicing and angel applied to structured data is investigated in [85, 86]. In the case of unstructured data, both techniques are applicable to high-dimensional transactional data as indicated in [85,86,127]. With the evolution of these techniques, there are various domains including social networks, wireless sensor network, and much more in that their application can be investigated. Differential privacy is applied to structured data for statistical analysis and privacy of relational data in [87,88,107]. It is also applied to unstructured data in [103, 108]. For

transactional data, it is claimed in [103], that this approach maintains scalability and is also applicable to the relational data, but the method used is limited to preserving privacy for count queries and frequent item sets.

3.3. Big Data application: An Analytical view in privacy perspective

We find that there exists a natural tendency of big data to be applied to e-health cloud scenario [122,123,152]. In Table 4 we present big data basics and its relation to e health, after that we describe privacy and scalability challenge posed by the integration of today's highly demanding paradigms of cloud computing, big data and privacy. The integration of today's highly demanding paradigms of cloud computing and big data create scalability and serious privacy issues. We can use cloud computing as a basic framework for big data systems to achieve some of its prime requirements like cost effectiveness, elasticity, and scalability. For big data management, more advanced data stores like NoSQL, Map Reduce, Hadoop etc. Exists to provide a scalable solution to handle big data challenges [126]. In the presence of big data, cloud computing privacy concerns become worse for organizations and users who want to save their private data at the cloud. As traditional data anonymization techniques mostly work on structured data of low volume, it is obvious that the techniques are inefficient to handle complex, variable, and the huge volume of unstructured/semi-structured data [98]. In literature, there exist some work that has been performed to solve the scalability and privacy concerns in terms of big data.

The focus here is to rearrange traditional privacy techniques to handle big data scalability and anonymization problem [98,124,125]. In [98], big data application in the cloud and anonymization scalability concerns are investigated. A specialization in anonymization process is performed in a way to use the full parallel capability of map reduce on the cloud. BUG is used and its variant in form of scalable advanced bottom up generalization is proposed. Anonymization is performed on split groups of smaller datasets in parallel. These intermediate outcomes are joined and again anonymization is applied to achieve data sets that satisfy *k*-anonymity. A hybrid scheme has been proposed which combines TDS and BUG.

Differential privacy provides strong theoretical privacy guarantees as compared to other techniques even in the presence of auxiliary information. In [15], Differential privacy, generalization, and personalized privacy[119] are used to achieve privacy in big data on the cloud. In [105,106,149-151], authors show that it is possible to apply differential privacy to the graph data. It is claimed that edge differential privacy is more suitable for achieving privacy while node differential privacy is apt to achieve better utility. Differential privacy can also be utilized in big data applications [125].

The authors in [125] discuss the compatibility of big data and differential privacy by mapping three main characteristics to data privacy. In these volume means large numbers of records, so to protect the privacy less noise will be needed, for velocity - speed with that data is processed or incoming data- is unaffected as noise is added to the output. In terms of variety,

Table 4: A brief overview of Big Data in e-health scenario

Big Data	Big Data mapping in e-health Scenario	Big Data Sources	Big Data Storage	Big Data Technology	Big Data Logical Analysis
Extensive datasets, primarily in the characteristics of volume, velocity and/or variety that require a scalable architecture for efficient storage, manipulation, and analysis.	**Value** Clinically relevant data need longitudinal Data for studying patient's history. **Volume** High-output technologies need for continuous monitoring of patients' critical information e.g. pulse rate, respiratory rate, and blood pressure. **Velocity** For fast healthcare decision support need high-speed processing. There is also increase in Data generation rate by health care infrastructures. **Variety** Various sources to provide unstructured/ semi-structured and streaming data. **Veracity** Data quality is not reliable as Data coming from uncontrolled environments. **Variability** Seasonal health consequences in patients and disease progression.	Heterogeneous sources providing unstructured/ semi structured and streaming data	Public/Private/ Hybrid Cloud	NoSQL data stores, Apache Hadoop and map/reduce	Need for real time analytics

all types of data (structured/semi-structured/unstructured) can be applied with differential privacy. Differential privacy provides robust privacy as compared to other techniques even in presence of auxiliary information moreover, it also improves privacy and utility balance. As research in this direction is in its evolutionary phase, it remains a challenge for existing privacy models and their underlying anonymization techniques to solve serious privacy concerns in E-health Cloud.

3.4. Utilization: privacy models and privacy preserving techniques

Different privacy models incorporate privacy preserving techniques to be effective against privacy threats. The purpose of this study is to find the privacy preserving techniques that are used in privacy models proposed to date. Moreover, it can be used to identify the techniques that can be proven to be the better candidate to use in previous privacy models like *k*-anonymity, *l*-diversity, *t*-closeness, and *m*-invariance etc. Table 5 depicts a comprehensive overview of privacy models in terms of underlying privacy technique moreover, it shows the achieved privacy level and utility. Privacy level is evaluated on the basis of privacy

preserving technique used in model against main privacy threats [14,128]. Utility is measured in terms of information loss which is an evaluation metrics as given in [155].

Discussion:

The research in data anonymization focuses on answering two fundamental questions: (*a) What Privacy-Aware Anonymity Technique should be adopted to transform the microdata to its noisy version (b) What privacy model should we judge i.e., whether the noisy data sufficiently protects sensitive information or not.* In Table 5, most privacy models use generalization and suppression to achieve privacy. Their excessive use causes reduced data utilization although required privacy level is achieved. The users of microdata are not able to achieve satisfactory results as generalized and suppressed data may produce incorrect results for query posed by different users. Also, these techniques provide minimum support against background knowledge attacks. For high dimensional data, both these techniques are non-responsive
because generalization of a large number of attributes can cause high information loss. It can also be seen that Slicing, Anatomy, and Angel are applied to *l*-diversity only.

Privacy Models										**Privacy level**			**Data Utility**
	Permutation	**Perturbation**	**Condensation**	**Randomization**	**Generalization**	**Suppression**	**Slicing**	**Anatomy**	**Angel**	**Identity Disclosure**	**Attribute Disclosure**	**Membership Disclosure**	
k-anonymity [112]	×	×	×	×	✓	✓	×	×	×	✓	×	×	Med.info. loss
l-Diversity [113]	×	×	×	×	✓	✓	✓	✓	✓	✓	✓	×	Med.info. loss
t-Closeness [113]	×	×	×	×	✓	✓	×	×	×	×	✓	×	Low.info. loss
(a, k) Anonymity [114]	×	×	×	×	✓	✓	×	×	×	×	✓	×	Med.info. loss
p-Sensitive k-anonymity [51]	×	×	×	×	✓	×	×	×	×	✓	×	×	Med.info. loss
(k, e) Anonymity Permutation [115]	✓	×	×	×	×	×	×	×	×	✓	✓	×	Low. info. loss
d-Presence [116]	×	×	×	×	✓	×	×	×	×	×	×	✓	Med.info. loss
m-invariance [117]	×	×	×	×	✓	×	×	×	×	×	✓	×	Med.info. loss
Personalized Privacy [118]	×	×	×	×	✓	×	×	×	×	×	×	✓	Med.info. loss
Differential Privacy [88]	×	✓	×	✓	✓	×	×	×	×	×	×	✓	High.info. loss
Extended *k*-anonymity model [154]	×	×	×	×	✓	×	×	×	×	×	✓	×	Med.info. loss
(ε, m)-Anonymity [156]	×	×	×	×	✓	×	×	×	×	×	✓	×	Med.info. loss
p^+-sensitive *k*-anonymity [55]	×	×	×	×	✓	×	×	×	×	✓	✓	×	Med.info. loss

In the case of slicing, we can say that it is partially applied to *k*-anonymity [61, 95,121]. Some of the statistical disclosure techniques like perturbation, randomization, and condensation can also be used in combination with different privacy models like *k*-

anonymity, *l*-diversity, and *t*-closeness. (k, e)-Anonymity uses permutation while Differential privacy can be attained by using randomization, perturbation or generalization. Angel combines anatomy and generalization to achieve *l*-diversity. The question that arises is: "*Why do we use generalization if it reduces data utilization".* Firstly, it preserves truthfulness of data and secondly, when used for marginal publication, generalization produces more accurate results than its counterparts. Other privacy techniques like permutation, perturbation, randomization, and condensation can be used with other privacy models depending upon the preferences of data publisher i.e., which privacy requirement (Figure 5) is of more interest to the data publisher. Similarly, we can use condensation in achieving *k*-anonymity and *l*-diversity where the level of privacy depends upon the value of *k or l* respectively. The abovementioned combination, in the context of data utilization, can be tailored accordingly if we condense large group of data into optimal size groups so that data statistics will not be lost. Privacy models can also be combined to achieve the best characteristics of models e.g. differential privacy and k-anonymity with the help of micro-aggregation is combined to enhance data utility [136].

In the case of dynamic data publication, *m*-invariance [118] and τ-safety [134] use generalization to achieve the given privacy requirements. The generalization-based *m*-invariance and τ-safety cause high information loss [128]. Since anatomy is shown to achieve better utility than generalization [74], it will be very interesting to investigate the anatomy with *m*-invariance and τ-safety.

From the above discussion, we emphasize that to manage the expanding requirements from data receivers, some privacy models and techniques have been proposed which consider certain scenarios against the structure of underlying data types, the possibilities of privacy threats inferences, etc. If we are looking to achieve an enhanced balance between privacy and utility, the answer lies in achieving the best combination of privacy preserving model and techniques.

This section presents a detailed account of privacy preserving anonymity techniques, their basic working mechanism, along with their merits and demerits. The choice of a privacy technique for a given data type is also a very important decision to be made by a data publisher. In this section, we also attempted to answer that "*how different privacy techniques respond to different data types?"* This will help the data publishers in selecting the appropriate privacy technique for their data.
In next section, we present privacy preserving requirements in cloud-based EHRs in detail with the help of taxonomy. Finally, we will evaluate each privacy technique based on the privacy requirements specific to the cloud-based EHRs.

4. Privacy preserving requirements for cloud- based EHR

This section will provide a complete review of the privacy-preserving requirements for EHRs based on their priority in the e-health cloud. When EHRs data is transferred to the cloud, the EHR system must set a few guarantees to save patients' delicate data alongside those of cloud

frameworks. The mix of these security prerequisites with those of the cloud will ensure the protection and security of outsourced EHRs [23].
Privacy requirements for cloud-based EHRs ensure that patients' EHRs remain secure from security and privacy breaches while at the same time its utility remains at the required level. We can say that these requirements act as a filter for the privacy techniques that can help a data publisher find the most appropriate technique for their data. Furthermore, as discussed before, EHRs in the cloud-based environment are susceptible to added privacy risks. These requirements ensure that the best balance of privacy and utility can be obtained for the EHRs in such a distributed environment. We divide privacy preserving requirements into three categories in the e-health cloud environment. The taxonomy is given in Figure 3.

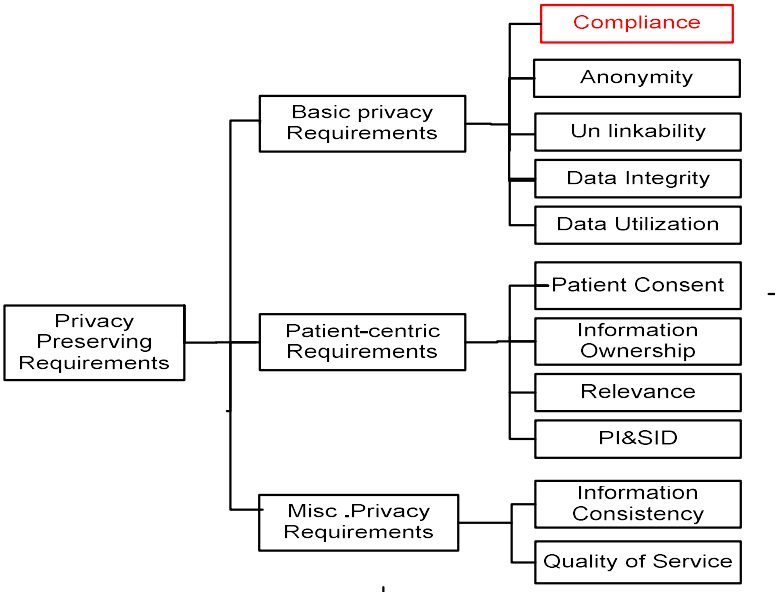

Figure 3: Taxonomy of Privacy Preserving Requirements

4.1 Basic privacy requirements

These are the mandatory privacy requirements in e-health cloud environment. Since the outsourced data are vulnerable to several other kinds of attacks, these basic requirements ensure that a privacy technique fulfills the basic guidelines to provide the optimal balance between privacy preservation and data utility.

4.1.1. Compliance: EHRs data has been continuously under consideration in various standards such as HIPAA, Open EHR, the HL7 and continuity of care document (CCD). HIPAA presents security standards and privacy assurance mechanisms to preserve health-related data. According to HIPAA personal identifiable information such as "*social security number, medical ID number, credit card number, driver's license number, home address, telephone number, medical records*", as protected health information (PHI). It was created to protect the individual's PHI. In 2009, HIPAA was updated into health information technology for economic and clinical health (HITECH). HITECH provides extra compliance standards in healthcare. The technical protection part of HIPAA defines the requirements that must be

satisfied in the design of access control, communication security for the advancement in healthcare [158].

Moreover, in recent years a new version of the EU Data Protection Directive the General data protection regulation GDRP has been introduced. According to the GDPR [9], the personal data is allowed only for processing, if the data owner has given its permission and most importantly, it must be limited to explicit data processing. The data controller is the entity that is accountable for data collection. It must specify a particular data processing purpose that cannot be modified later unreasonably. In the case of GDPR, data processing should be done only on anonymized data with great care [11].

More effective regulations about various aspects of personal data like the treatment of confidential data, informed consent, data treatment and flow of personal data are described in detail in [9]. These features could be fully analysed only after the new regulations will effectively enter into force.

4.1.2 Anonymity (AN): Anonymity means that patients related health data is outsourced at the public cloud in the form that could be unidentified by CSPs, researchers and external adversaries. Patient's information, for example, identifiers (Name, social security information number, delicate health data) ought to stay avoided from data recipients. In anonymization process, patient's personal information is transformed into anonymized form by the data processor. Patient and Healthcare organization are fully responsible for consent purpose and it is mandatory. Data processor adopt necessary privacy measures like application of privacy preserving techniques to achieve anonymity before other data processing. Pseudo-anonymity is currently used in several approaches to achieve anonymity; however, true anonymity can only be achieved by using anonymization techniques.

4.1.3 Unlinkability (UN): Adversaries should not be able to link patients' identifying information and patients sensitive attribute information. Unlinkability can be properly achieved by several privacy preserving models (*k*-anonymity, t-closeness etc...) and techniques E.g. "Generalization and Suppression" etc.

4.1.4 Data Integrity: means that health data should remain accurate and consistent when it is stored in the cloud. Any prohibitive action of user's data should not be modified [24]. Privacy preserving techniques should preserve EHRs data at the same time EHRs data must show accurate content

4.1.5 Data utilization. Data utilization means that EHRs data remains useful after applying Standard Privacy rules like GDPR Compliance and other security and privacy protection measures [9]. Data should remain useful to other entities in health care domain even after applying the privacy protection measures with the help of Anonymization Techniques.

.

4.2 Patient Centric Privacy Requirements

These privacy requirements are directly related to patients. Some of the requirements like patient consent and information ownership purely depend on the information owner (patient). Some other requirements like relevance, patient's identifiable & sensitive information disclosure is bound to be addressed when we use privacy techniques to preserve patients' EHRs data.

4.2.1 Patient Consent: Patient's consent as per rules and legislation is the authority to permit or deny access to health information with exception to emergency situations. In the cloud, another important issue is to protect patients consent based access control. There may be a situation when a health professional wants a patient to provide a history of medical records

that are not stored earlier at EHRs hosted domain of healthcare organization. It requires the consent of both patient and authorization from their respective healthcare organization.

4.2.2 Information Ownership: Information Ownership is mandatory to protect Patient's EHR information from illegal access and perversion of patient's sensitive information. The creator of specific information is generally believed to be the owner of that information but unfortunately, the ownership of EHR is not completely related to the patients, though; they have full right to access their medical information. The medical personnel, creating and storing that information, are also termed as owners of EHR [20, 24].

4.2.3 Relevance: It means that only relevant person e.g. medical personnel are involved in the process of diagnosis and treatment should have the authority to access the EHRs data. Default permission mechanism is preferred in this regard [24]. Relevance can be achieved by incorporating privacy anonymization techniques that have the flexibility to give only that portion of Patients data that is relevant to the requesting authority.

4.2.4 Patient's identifiable & sensitive information disclosure (PI&SID): It incorporates any information that can be utilized to distinguish an individual, for example, name, address, and so forth. Patients' identifiable information can be associated with other data to distinguish or find individuals, for example, social relations data.

4.2.5 Sensitive Information and Disclosure: The EHRs data is also considered as sensitive personal information under GDPR Directive and falls in "special categories of data". It also specifically includes *"genetic data and biometric data"* where processed "*to uniquely identify a person*" [157]. Sensitive information requires extra security measures. It also includes information about religion, race, or well-being. Some other information may also be viewed as sensitive such as private fiscal information, job performance report and personally identifiable information e.g. biometric information [2].

4.3 Cloud Specific Privacy Requirements

Privacy requirements like information consistency and quality of service relate to the fundamental aspects of information security. By following these requirements, a privacy technique can further enhance the privacy vs. utility trade-off in the outsourced data.

4.3.1 Information Consistency (IC)

Different versions of patients EHRs exist for different personnel like health professionals. Health providers, employees of cloud service providers etc. In such a distributed environment, all versions of EHRs must be consistent with a change in the information that could occur after EHR update.

To maintain IC, interoperability in EHRs related to different health care organizations must be carefully monitored i.e., a revision warning mechanism must be performed to show changes to the EHRs data. This process must enable access to the prior versions of the EHRs if required, for information consistency [24]. Moreover, in the context of privacy, after applying anonymization technique, data in EHRs must be consistent.

4.3.2 Quality of Service (QoS)

Medicinal records (EHRs) holding patient data may belong to various health organizations. EHRs ought to be accessible anytime and at anyplace, permitting healthcare providers to provide proper patients healthcare investigation and medical treatment [2].

The quality of Service (QoS) of medical records directly depends upon medical device used for diagnosis purpose in health institutions. It includes how the device is being used by different medical staff, device model and calibration data used for taking readings during the

diagnosis process. By improving all the parameter will improve the overall QoS of EHRs [137].

In this section, we briefly reviewed the privacy requirements like Anonymity, Data integrity, Data utilization, Relevance etc. keeping in mind the application of privacy preserving techniques. Some patient-centric requirements like patient consent and information ownership are not affected by privacy preserving techniques as these requirements directly relate to patients or information owner while the other requirements need to be effectively addressed by privacy techniques to achieve the best trade-off between privacy and utility for outsourced data. We have shortly described the GDPR [157] in the EHRs context from a legal point of view. However, there is an imperative need to re-investigate the EHRs privacy requirements in more details in the GDPR perspective. In the next section, we highlight the relation between the privacy techniques and privacy requirements.

5. Comparison: privacy-aware anonymity-based techniques against privacy preserving requirements.

In this section, we perform a comparative analysis of privacy techniques based on the privacy preserving requirements. Table 6 depicts the privacy requirements as seen by different privacy techniques.

In Table 6, it can be observed that we are not clear about the patient-centric requirements like patient's consent (PC) and information ownership (IO). The reason is that these are purely E-health and EHR related ones and require patient's control and feedback. However, anatomy and angel are technically shown to satisfy relevance (RL), all privacy aware anonymity-based techniques' main objective is to anonymize the microdata, so in a broader view, all the techniques satisfy anonymity (AN), unlikability (UN), patient's identifying information (PID) and patient's sensitive information disclosure (PSD).

Privacy level is added to the table to highlight its relationship with data utilization. For all privacy aware techniques, the general trend is that where privacy is satisfied, data utilization is low. In some cases, like suppression, privacy level is high, but data utilization is not satisfied, as suppression deletes the data value. In privacy techniques like permutation, perturbation, condensation, randomization and differential privacy, data are modified in different ways, so the data utilization is low. Generalization is shown to achieve better privacy and data utilization. In privacy techniques like Anatomy, Slicing, and Angel, there is a general trend that highlights high privacy level and high data utilization.

Table 6: Comparative analysis of privacy-preserving techniques

Anonymity based Privacy Techniques	**Privacy Requirements**												
	P	**DI**	**AN**	**UN**	**PID**	**PSD**	**DU**	**PC**	**RL**	**IO**	**IC**	**QoS**	**CMP**
Permutation [71,76]	√	√	√	√	√	√	↓	-	-	-	×	×	√
Perturbation [73-77]	√	×	√	√	√	√	↓	-	-	-	×	×	√
Condensation [73,77,80]	√	×	√	√	√	√	↓	-	-	-	×	×	√
Randomization [74-	√	×	√	√	√	√	↓	-	-	-	×	×	√

79]													
Generalization [73,82]	√	√	√	√	√	√	↑	-	-	-	↓	↓	√
Suppression [73,81,82]	↑	×	√	√	√	√	×	-	-	-	×	×	√
Anatomy [84]	√	↑	√	√	√	√	↑	-	√	-	↑	↑	√
Angel [86]	↑	√	√	√	√	√	↑	-	√	-	↓	↑	√
Slicing [85]	↑	√	√	√	√	√	↑	-	-	-	×	↑	√
Differential privacy [88]	↑	√	√	√	√	√	↓	-	-	-	×	↓	√

Symbols used for privacy requirements: √: Satisfied, ×: Not satisfied, ↑: High, ↓: Low, -: Not known

P: Privacy level, **DI**: Data Integrity**, AN**: Anonymity, **UN**: Unlinkability, **PID**: Patient's identifying information disclosure, **PSD**: Patient's sensitive data disclosure, DU: Data Utilization, **PC**: Patient's consent, **RL**: Relevance, **IO**: Information ownership, **IC**: Information Consistency, **QoS**: Quality of Service, **CMP: Compliance.**

*In the context of E-health cloud privacy requirements throughout discussion when using the term "Data" it will directly have applied to EHRs data.

Data Integrity (DI) is affected when a random value is added or multiplied to numerical attribute values in microdata table (EHRs Data). Consequently, data integrity is lost or not satisfied in the case of perturbation, randomization, and condensation and differential privacy. Permutation rearranges numerical attribute values thus; it does not distort data integrity.

Data integrity is satisfied in the case of generalization, angel, slicing and differential privacy but it is not satisfied in the case of suppression.

In all privacy techniques, information consistency (IC) is not up to the mark except anatomy, because all attribute values (QID, SA) remain unmodified. However, in the case of Generalization and Angel anonymize data show low information consistency. There is no Information Consistency (IC), in the case of all privacy techniques that distort the internal distribution of attributes values (QID, SA). e.g. permutation, perturbation, differential privacy etc.

The quality of Service (QoS) is either not satisfied or is low in permutation. Perturbation, condensation, randomization, generalization, suppression and differential privacy; QoS is high in case of anatomy, slicing or angel. Privacy techniques here will affect (QoS) in terms of data quality (EHRs data). Specific readings taken during diagnosis process of a disease will be greatly modified by applying certain privacy techniques e.g. permutation, suppression, differential privacy etc., So in all these cases, we avoid using such privacy techniques to e-health data. anatomy, slicing, and angel keep the quality of data to certain acceptable level. The compliance under GDPR directive, satisfied by all privacy preserving techniques. As it is the responsibility of data controller (EHRs holder health organization) and data processor to adopt necessary privacy and security measures.

From the discussion above, we can conclude that each privacy technique has its merits and demerits regarding privacy requirements. Its main aim is to get some knowledge about the privacy techniques that can be adapted in e-health. Privacy techniques like anatomy, angel,

slicing and differential privacy are the most prominent techniques that satisfied maximum privacy aware requirements that are e-health specific. However, about differential privacy, there are some reservations as the technique modifies the numerical values by adding some noise in patients' data and not applicable to non-numerical attributes (categorical attributes) of patients EHR. Also, for above-mentioned techniques, there is a general trend that shows high privacy level and data utilization. In general, choice of a given privacy technique depends on the necessities of the data holder keeping in mind the trade-off among different privacy requirements as one requirement that is important to one data holder might not be important to the other.

6. Conclusion and future work

The privacy of publicly released data is a challenging task. A plethora of work is performed to provide maintain the privacy of the data. In the said perspective, we conducted a comprehensive research to do an in-depth analysis of privacy preserving techniques in e-health cloud. First, an effort is made to determine *what is the actual need for preserving the privacy of EHRs, stored at the cloud?* The main theme of our work is to highlight the need for *privacy preserving techniques when we outsource EHRs data to the cloud.* We presented privacy techniques with their merits, demerits, along with their applicability to the taxonomy of various data types. We tried to fill the gap in choosing the best combination of privacy techniques and privacy models, which in turn, can substantially improve the privacy level and the utility of released data. Finally, we presented EHRs specific privacy requirements in a separate taxonomy depending upon their priority in e-health cloud. Moreover, an updated privacy preserving analysis in advanced version of EU Directive GDRP is also presented. We provided a deep comparative analysis of privacy techniques based on the privacy preserving requirements.

As future work, we are keen on exploring the most innovative ideas and features in privacy-preserving techniques and models in terms of EHRs data confidentiality in cloud scenarios. Another area that needs considerable attention is the use of contemporary Access Control mechanisms to achieve fine-grained access levels with privacy protection techniques. Privacy disclosures identification and prevention for cloud based EHRs in real life dataset scenarios also need adequate investigation.

References

1. AbuKhousa E, Mohamed N, Al-Jaroodi J. e-Health cloud: opportunities and challenges. Future Internet. 2012 Jul 4;4(3):621-45.
2. Lynda, Kacha, Oukid-KhouasSaliha, and BenblidiaNadjia. "Data security and privacy in E-health Cloud: Comparative study." *Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication*. ACM, 2015.

3. Löhr, Hans, Ahmad-Reza Sadeghi, and Marcel Winandy. "Securing the e-health cloud." *Proceedings of the 1st ACM International Health Informatics Symposium*. ACM, 2010.
4. Bahgaa Arshdeep, and Vijay K. Madisetti. "A cloud-based approach for interoperable electronic health records (EHRs)." *IEEE Journal of Biomedical and Health Informatics* 17.5 (2013): 894-906.
5. (2012). VistA Monograph [Online]. Available: www.va.gov/vista monograph
6. OpenEHR. (2012). [Online]. Available: http://www.openehr.org
7. Available: http://www.athenahealth.com
8. Achampong, Emmanuel Kusi. "Electronic Health Record (EHR) and Cloud Security: The Current Issues." *International Journal of Cloud Computing and Services Science* 2.6 (2013): 417.
9. EU Commission website. https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en. Accessed 28 May 2019.
10. Demotes-Mainard, Jacques, et al. "How the new European data protection regulation affects clinical research and recommendations?" *Thérapie* 74.1 (2019): 31-42.
11. Gruschka, Nils, et al. "Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR." *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018.
12. Abbas, Assad, and Samee U. Khan. "e-Health Cloud: Privacy Concerns and Mitigation Strategies." *Medical Data Privacy Handbook*. Springer International Publishing, 2015. 389-421
13. Taneja, Himanshu, and Ashutosh Kumar Singh. "Preserving Privacy of Patients Based on Re-identification Risk." *Procedia Computer Science* 70 (2015): 448-454.
14. Gkoulalas-Divanis, Aris, GrigoriosLoukides, and Jimeng Sun. "Publishing data from electronic health records while preserving privacy: a survey of algorithms." *Journal of biomedical informatics* 50 (2014): 4-19.
15. Wei Wang, Lei Chen, Qian Zhang, Outsourcing high-dimensional healthcare data to cloud with personalized privacy preservation, Computer Networks, Volume 88, 9 September 2015, Pages 136-148, ISSN 1389-1286, http://dx.doi.org/10.1016/j.comnet.2015.06.014
16. Takabi, Hassan. "Privacy aware access control for data sharing in cloud computing environments." *Proceedings of the 2nd international workshop on Security in cloud computing*. ACM, 2014
17. Omnibus, Hipaa rule in the Federal Register, 2013, http://www.gpo.gov/fdsys/pkg/FR-2013-01-25/pdf/2013-01073.pdf.
18. W. Wang, Q. Zhang Towards long-term privacy preservation: A context aware perspective, IEEE Wireless Commun. (2015)
19. Pandilakshmi, K. R., and G. RashithaBanu. "An Advanced Bottom up Generalization Approach for Big Data on Cloud." *Volume* 3 (2014): 1054-1059
20. Heurix, Johannes, et al. "A taxonomy for privacy enhancing technologies." *Computers & Security* 53 (2015): 1-17
21. Pandilakshmi, K. R., and G. RashithaBanu. "An Advanced Bottom up Generalization Approach for Big Data on Cloud." *Volume* 3 (2014): 1054-1059.

22. Sedayao, Jeff. "Enhancing cloud security using data anonymization." *White Paper, Intel Coporation* (2012).
23. Sinha,Tanmay, et al. "Trends and research directions for privacy preserving approaches on the cloud." *Proceedings of the 6th ACM India Computing Convention*. ACM, 2013.
24. Rodrigues, Joel JPC, et al. "Analysis of the security and privacy requirements of cloud-based electronic health records systems." *Journal of medical Internet research* 15.8 (2013): e186.
25. Abbas, Asad, and Samee U. Khan. "A review on the state-of-the-art privacy-preserving approaches in the e-health clouds." *Biomedical and Health Informatics, IEEE Journal of* 18.4 (2014): 1431-1441
26. Sahai, Amit, and Brent Waters. "Fuzzy identity-based encryption." *Advances in Cryptology–EUROCRYPT 2005*. Springer Berlin Heidelberg, 2005. 457-473.
27. Goyal, Vipul, et al. "Attribute-based encryption for fine-grained access control of encrypted data." *Proceedings of the 13th ACM conference on Computer and communications security*. Acm, 2006
28. Bethencourt, John, Amit Sahai, and Brent Waters. "Ciphertext-policy attribute-based encryption." *Security and Privacy, 2007. SP'07. IEEE Symposium on*. IEEE, 2007.
29. Chase, Melissa, and Sherman SM Chow. "Improving privacy and security in multi-authority attribute-based encryption." *Proceedings of the 16th ACM conference on Computer and communications security*. ACM, 2009
30. Danwei, Chen, et al. "Securing patient-centric personal health records sharing system in cloud computing." *Communications, China* 11.13 (2014): 121-127.
31. Song, Dawn Xiaodong, David Wagner, and Adrian Perrig. "Practical techniques for searches on encrypted data." *Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on*. IEEE, 2000.
32. Narayan, Shivaramakrishnan, Martin Gagné, and ReihanehSafavi-Naini. "Privacy preserving EHR system using attribute-based infrastructure."*Proceedings of the 2010 ACM workshop on Cloud computing security workshop*. ACM, 2010.
33. Shamir, Adi. "Identity-based cryptosystems and signature schemes."*Advances in cryptology*. Springer Berlin Heidelberg, 1984.
34. Benaloh, Josh, et al. "Patient controlled encryption: ensuring privacy of electronic medical records." *Proceedings of the 2009 ACM workshop on Cloud computing security*. ACM, 2009.
35. Gentry, Craig. "Fully homomorphic encryption using ideal lattices." *STOC*. Vol. 9. 2009.
36. Naehrig, Michael, Kristin Lauter, and Vinod Vaikuntanathan. "Can homomorphic encryption be practical?" *Proceedings of the 3rd ACM workshop on Cloud computing security workshop*. ACM, 2011
37. Lin, Huang, et al. "CAM: cloud-assisted privacy preserving mobile health monitoring." *Information Forensics and Security, IEEE Transactions on* 8.6 (2013): 985-997.
38. Chen, Yu-Yi, Jun-Chao Lu, and Jinn-Ke Jan. "A secure EHR system based on hybrid clouds." *Journal of medical systems* 36.5 (2012): 3375-3384.

39. Li, Zhuo-Rong, et al. "A secure electronic medical record sharing mechanism in the cloud computing platform." *Consumer Electronics (ISCE), 2011 IEEE 15th International Symposium on*. IEEE, 2011.
40. Kumar, Naveen, Anish Mathuria, and Manik Lal Das. "Achieving forward secrecy and unlinkability in cloud-based personal health record system." *2015 IEEE Trustcom/BigDataSE/ISPA*. Vol. 1. IEEE, 2015.
41. Zhang, Rui, and Ling Liu. "Security models and requirements for healthcare application clouds." *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*. IEEE, 2010.
42. Narayan, Shivaramakrishnan, Martin Gagné, and ReihanehSafavi-Naini. "Privacy preserving EHR system using attribute-based infrastructure."*Proceedings of the 2010 ACM workshop on Cloud computing security workshop*. ACM, 2010.
43. Alshehri, Suhair, StanisławRadziszowski, and Rajendra K. Raj. "Designing a secure cloud-based ehr system using ciphertext-policy attribute-based encryption." *Proceedings of the Data Management in the Cloud Workshop, Washington, DC, USA*. 2012.
44. Barua, Mrinmoy, et al. "ESPAC: Enabling Security and Patient-centric Access Control for eHealth in cloud computing." *International Journal of Security and Networks* 6.2-3 (2011): 67-76
45. Takabi, Hassan. "Privacy aware access control for data sharing in cloud computing environments." *Proceedings of the 2nd international workshop on Security in cloud computing*. ACM, 2014.
46. Peleg, Mor, et al. "Situation-based access control: Privacy management via modelling of patient data access scenarios." *Journal of biomedical informatics* 41.6 (2008): 1028-1040
47. Haas, Sebastian, et al. "Aspects of privacy for electronic health records."*International journal of medical informatics* 80.2 (2011): e26-e31.
48. Xu, Liangyu, Armin B. Cremers, and Tobias Wilken. "Pseudonymization for secondary use of cloud based electronic health records." (2015).
49. Pecarina,John, Shi Pu, and Jyh-Charn Liu. "SAPPHIRE: Anonymity for enhanced control and private collaboration in healthcare clouds." *Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on*. IEEE, 2012.
50. Tong, Yue, et al. "Cloud-assisted mobile-access of health data with privacy and auditability." *Biomedical and Health Informatics, IEEE Journal of* 18.2 (2014): 419-429.
51. Riedl B, Grascher V, Fenz S, Neubauer T. Pseudonymization for improving the privacy in e-health applications. In: Proc annual Hawaii intconf system sciences; 2008. p. 1–9.
52. Huang LC, Chu HC, Lien CY, Hsiao CH, Kao T. Privacy preservation and information security protection for patients' portable electronic health records.ComputBiol Med 2009;39(9):743–50.

53. Alhaqbani B, Fidge C. Privacy-preserving electronic health record linkage using pseudonym identifiers. In: Proc intconf e-health networking,applications and services healthcom; 2008. p. 108–17.
54. Yang, Ji-Jiang, Jian-Qiang Li, and Yu Niu. "A hybrid solution for privacy preserving medical data sharing in the cloud environment." *Future Generation Computer Systems* 43 (2015): 74-86.
55. X. Sun, L. Sun, and H. Wang, "Extended k-anonymity models against sensitive attribute disclosure," *Comput. Commun.*, vol. 34, no. 4, pp. 526–535, 2011.
56. El Emam, Khaled, and Fida Kamal Dankar. "Protecting privacy using k-anonymity." *Journal of the American Medical Informatics Association* 15.5 (2008): 627-637.
57. Gionis, Aristides, ArnonMazza, and TamirTassa. "k-Anonymization revisited." *2008 IEEE 24th International Conference on Data Engineering*. IEEE, 2008.
58. W Bahgaa Arshdeep, and Vijay K. Madisetti. "A cloud-based approach for interoperable electronic health records (EHRs)." IEEE Journal of Biomedical and Health Informatics 17.5 (2013): 894-906.
59. Truta, Traian Marius, and Bindu Vinay. "Privacy Protection: p-Sensitive k-Anonymity Property." *ICDE workshops*. 2006.
60. Nergiz, Mehmet Ercan, and Chris Clifton. "δ-presence without complete world knowledge." *IEEE Transactions on Knowledge and Data Engineering*22.6 (2010): 868-883.
61. Samarati, Pierangela. "Protecting respondents identities in microdata release." *IEEE transactions on Knowledge and Data Engineering* 13.6 (2001): 1010-1027.
62. LeFevre, Kristen, David J. DeWitt, and Raghu Ramakrishnan. "Incognito: Efficient full-domain k-anonymity." *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*. ACM, 2005.
63. LeFevre, Kristen, David J. DeWitt, and Raghu Ramakrishnan. "Mondrian multidimensional k-anonymity." *22nd International Conference on Data Engineering (ICDE'06)*. IEEE, 2006.
64. LeFevre, Kristen, David J. DeWitt, and Raghu Ramakrishnan. "Workload-aware a Jian, et al. "Utility-based anonymization using local recoding."*Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006.
65. Xu, Jian, et al. "Utility-based anonymization using local recoding."*Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006.
66. Fung, Benjamin CM, Ke Wang, and Philip S. Yu. "Top-down specialization for information and privacy preservation." *21st International Conference on Data Engineering (ICDE'05)*. IEEE, 2005.
67. Li, Jiuyong, et al. "Achieving k-anonymity by clustering in attribute hierarchical structures." *International Conference on Data Warehousing and Knowledge Discovery*. Springer Berlin Heidelberg, 2006.

68. Wong, Raymond Chi-Wing, et al. "(α, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing." *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006.
69. Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian. "Closeness: A new privacy measure for data publishing." *IEEE Transactions on Knowledge and Data Engineering* 22.7 (2010): 943-956.
70. Nergiz, Mehmet Ercan, and Chris Clifton. "δ-presence without complete world knowledge." *IEEE Transactions on Knowledge and Data Engineering*22.6 (2010): 868-883.
71. Burke, Mark John. "Enabling anonymous crime reporting on mobile phones in the developing world." (2013).
72. Panackal, Jisha Jose, and Anitha S. Pillai. "Privacy Preserving Data Mining: An Extensive Survey." *ACEEE. International Conference on Multimedia Processing, communication and Information Technology*. 2013.
73. Fung, Benjamin CM. *Privacy-preserving data publishing*. Diss. Simon Fraser University, 2007.
74. Sehatkar, Morvarid. *Towards a Privacy Preserving Framework for Publishing Longitudinal Data*. Diss. University of Ottawa, 2014.
75. Gkountouna, Olga. *A Survey on Privacy Preservation Methods*. NTUA, Technical Report, 2011.
76. Panackal, Jisha Jose, and Anitha S. Pillai. "Privacy Preserving Data Mining: An Extensive Survey." *ACEEE. International Conference on Multimedia Processing, communication and Information Technology*. 2013.
77. Aggarwal, Charu C., and S. Yu Philip. *A general survey of privacy-preserving data mining models and algorithms*. Springer US, 2008.
78. Guo, Ling. *Randomization Based Privacy Preserving Categorical Data Analysis*. Diss. The University of North Carolina at Charlotte, 2010
79. Kiran, P., and N. P. Kavya. "A Survey on Methods, Attacks and Metric for Privacy Preserving Data Publishing." *International Journal of Computer Applications* 53.18 (2012): 20-28.
80. Aggarwal, Charu C., and S. Yu Philip. "A condensation approach to privacy preserving data mining." *Advances in Database Technology-EDBT 2004*. Springer Berlin Heidelberg, 2004. 183-199.
81. Xu, Yang, et al. "A survey of privacy preserving data publishing using generalization and suppression." *Appl. Math* 8.3 (2014): 1103-1116.
82. Sweeney, Latanya. "Achieving k-anonymity privacy protection using generalization and suppression." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.05 (2002): 571-588.
83. Zigomitros, Athanasios, AgustiSolanas, and ConstantinosPatsakis. "The role of inference in the anonymization of medical records." *Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on*. IEEE, 2014.
84. Xiao, Xiaokui, and Yufei Tao. "Anatomy: Simple and effective privacy preservation." *Proceedings of the 32nd international conference on Very large data bases*. VLDB Endowment, 2006.

85. Li, Tiancheng, et al. "Slicing: A new approach for privacy preserving data publishing." *Knowledge and Data Engineering, IEEE Transactions on* 24.3 (2012): 561-574
86. Tao, Yufei, et al. "Angel: Enhancing the utility of generalization for privacy preserving publication." *Knowledge and Data Engineering, IEEE Transactions on* 21.7 (2009): 1073-1087
87. Brickell, Justin Lee. "Privacy-preserving computatiaon for data mining." (2009).
88. Dwork, Cynthia. "Differential privacy: A survey of results." *Theory and applications of models of computation*. Springer Berlin Heidelberg, 2008. 1-19.
89. Ghinita, Gabriel, Yufei Tao, and PanosKalnis. "On the anonymization of sparse high-dimensional data." *2008 IEEE 24th International Conference on Data Engineering*. Ieee, 2008.
90. Zhang, Qing, et al. "Aggregate query answering on anonymized tables." *2007 IEEE 23rd International Conference on Data Engineering*. IEEE, 2007.
91. Zheleva, Elena, and LiseGetoor. "Preserving the privacy of sensitive relationships in graph data." *Privacy, security, and trust in KDD*. Springer Berlin Heidelberg, 2008. 153-171.
92. Saygin, Yücel, D. Hakkani-Tur, and Gökhan Tur. "Sanitization and anonymization of document repositories." *Web and information security*(2006): 133.
93. Li, Chun, Charu C. Aggarwal, and Jianyong Wang. "On Anonymization of Multi-graphs." *SDM*. 2011.
94. Poulis, Giorgos, et al. "Anonymizing data with relational and transaction attributes." *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer Berlin Heidelberg, 2013.
95. Terrovitis, Manolis, Nikos Mamoulis, and PanosKalnis. "Privacy-preserving anonymization of set-valued data." *Proceedings of the VLDB Endowment* 1.1 (2008): 115-125.
96. LIU, Junqiang. "Optimal Anonymization for Transaction Publishing." *Chinese Journal of Electronics* 20.2 (2011).
97. Jiang, Wei, et al. "t-Plausibility: semantic preserving text sanitization."*Computational Science and Engineering, 2009. CSE'09. International Conference on*. Vol. 3. IEEE, 2009.
98. Pandilakshmi, K. R., and G. RashithaBanu. "An Advanced Bottom up Generalization Approach for Big Data on Cloud." *Volume* 3 (2014): 1054-1059.
99. Liu, Junqiang, and Ke Wang. "Anonymizing transaction data by integrating suppression and generalization." *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg, 2010.
100. Chakaravarthy, Venkatesan T.,et al. "Efficient techniques for document sanitization." *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 2008.
101. Chen, Rui, et al. "Privacy-preserving trajectory data publishing by local suppression." *Information Sciences* 231 (2013): 83-97.

102. Xu, Yabo, et al. "Anonymizing transaction databases for publication."*Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2008.
103. JChen, Rui, et al. "Publishing set-valued data via differential privacy."*Proceedings of the VLDB Endowment* 4.11 (2011): 1087-1098.
104. Shrivastva, Krishna Mohan Pd, M. A. Rizvi, and Shailendra Singh. "Big Data Privacy Based On Differential Privacy a Hope for Big Data." *Computational Intelligence and Communication Networks (CICN), 2014 International Conference on*. IEEE, 2014.
105. Gupta, Anupam, Aaron Roth, and Jonathan Ullman. "Iterative constructions and private data release." *Theory of Cryptography Conference*. Springer Berlin Heidelberg, 2012.
106. Sala, Alessandra, et al. "Sharing graphs using differentially private graph models." *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. ACM, 2011.
107. Li, Ninghui, WahbehQardaji, and Dong Su. "On sampling, anonymization, and differential privacy or, k-anonymization meets differential privacy."*Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*. ACM, 2012.
108. Andrés, Miguel E., et al. "Geo-indistinguishability: Differential privacy for location-based systems." *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 2013.
109. De Mauro, Andrea, Marco Greco, and Michele Grimaldi. "What is big data? A consensual definition and a review of key research topics." *AIP Conference Proceedings*. Vol. 1644. No. 1. 2015.
110. Sagiroglu, Seref, and DuyguSinanc. "Big data: A review." *Collaboration Technologies and Systems (CTS), 2013 International Conference on*. IEEE, 2013.
111. Gharehchopogh, FarhadSoleimanian, and ZeinabAbbasiKhalifelu. "Analysis and evaluation of unstructured data: text mining versus natural language processing." *Application of Information and Communication Technologies (AICT), 2011 5th International Conference on*. IEEE, 2011.
112. Sweeney, Latanya. "k-anonymity: A model for protecting privacy." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10.05 (2002): 557-570.
113. Machanavajjhala, Ashwin, et al. "l-diversity: Privacy beyond k-anonymity."*ACM Transactions on Knowledge Discovery from Data (TKDD)* 1.1 (2007): 3.
114. Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian. "t-closeness: Privacy beyond k-anonymity and l-diversity." *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*. IEEE, 2007.
115. Wong, Raymond Chi-Wing, et al. "(α, k)-anonymity: an enhanced k-anonymity model for privacy preserving data publishing." *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2006.
116. Zhang, Qing, et al. "Aggregate query answering on anonymized tables." *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*. IEEE, 2007.

117. Nergiz, Mehmet Ercan, Maurizio Atzori, and Chris Clifton. "Hiding the presence of individuals from shared databases." *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*. ACM, 2007
118. X. Xiao and Y. Tao. *m*-invariance: Towards privacy preserving re-publication of dynamic datasets. In *SIGMOD*, 2007
119. Xiao, Xiaokui, and Yufei Tao. "Personalized privacy preservation." *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*. ACM, 2006.
120. Kargupta, Hillol, et al. "Random-data perturbation techniques and privacy-preserving data mining." *Knowledge and Information Systems* 7.4 (2005): 387-414
121. Terrovitis, Manolis, et al. "Privacy preservation by disassociation. "*Proceedings of the VLDB Endowment* 5.10 (2012): 944-955.
122. De Mauro, Andrea, Marco Greco, and Michele Grimaldi. "What is big data? A consensual definition and a review of key research topics." *AIP Conference Proceedings*. Vol. 1644. No. 1. 2015.
123. Nair, Lekha R., and Sujala D. Shetty. "Research in Big Data and Analytics: An Overview." *International Journal of Computer Applications* 108.14 (2014).
124. Zhang, Xuyun, et al. "A hybrid approach for scalable sub-tree anonymization over big data using MapReduce on cloud." *Journal of Computer and System Sciences* 80.5 (2014): 1008-1020.
125. Shrivastva, Krishna Mohan Pd, M. A. Rizvi, and Shailendra Singh. "Big Data Privacy Based On Differential Privacy a Hope for Big Data." *Computational Intelligence and Communication Networks (CICN), 2014 International Conference on*. IEEE, 2014.
126. Hu, Han, et al. "Toward scalable systems for big data analytics: A technology tutorial." *IEEE Access* 2 (2014): 652-687.
127. Li, Dong, et al. "Permutation anonymization." *Journal of Intelligent Information Systems* (2015): 1-19
128. Fung, Benjamin, et al. "Privacy-preserving data publishing: A survey of recent developments." *ACM Computing Surveys (CSUR)* 42.4 (2010): 14.
129. N. Cao, C. Wang, M. Li, K. Ren, W. Lou."Privacy-preserving multi-keyword ranked search over encrypted cloud data"*Proceeding of the IEEE INFOCOM (2011)*
130. J. Yuan, S. Yu. "Efficient privacy-preserving biometric identification in cloud computing". *Proceedings of the IEEE INFOCOM (2013)*
131. C. Wang, K. Ren, S. Yu, K.M.R. Urs. "Achieving usable and privacy-assured similarity search over outsourced cloud data" Proceedings of the IEEE INFOCOM (2012)1
132. Zhang, Kehuan, et al. "Sedic: privacy-aware data intensive computing on hybrid clouds." Proceedings of the 18th ACM conference on Computer and communications security. ACM, 2011.
133. Zhou, Zhigang, et al. "Prometheus: Privacy-aware data retrieval on hybrid cloud." INFOCOM, 2013 Proceedings IEEE. IEEE, 2013.
134. Anjum, A., Raschia, G.: Anonymizing sequential releases under arbitrary updates. In: Proceedings of the Joint EDBT/ICDT 2013 Workshops, EDBT '13, pp. 145–154 (2013)

135. Ali, Mazhar, Samee U. Khan, and Athanasios V. Vasilakos. "Security in cloud computing: Opportunities and challenges." *Information Sciences* 305 (2015): 357-383.
136. Soria-Comas, Jordi, et al. "Enhancing data utility in differential privacy via micro aggregation-based k-anonymity." *The VLDB Journal* 23.5 (2014): 771-794.
137. Van Deursen, Ton, Paul Koster, and Milan Petkovic. "Reliable personal health records." *Studies in health technology and informatics* 136 (2008): 484.
138. Löhr, Hans, Ahmad-Reza Sadeghi, and Marcel Winandy. "Securing the e-health cloud." *Proceedings of the 1st ACM International Health Informatics Symposium*. ACM, 2010.
139. Pino, Carmelo, and Roberto Di Salvo. "A survey of cloud computing architecture and applications in health." *International Conference on Computer Science and Electronics Engineering*. 2013.
140. Chandrasekaran, Srimathi, Subaji Mohan, and Rajesh Natarajan. "Survey on HealthCloud characteristics." *Health and Technology* 5.2 (2015): 135-146.
141. Pussewalage, Harsha S. Gardiyawasam, and Vladimir A. Oleshchuk. "Privacy preserving mechanisms for enforcing security and privacy requirements in E-health solutions." *International Journal of Information Management* 36.6 (2016): 1161-1173.
142. Yüksel, Buket, AlptekinKüpçü, and ÖznurÖzkasap. "Research issues for privacy and security of electronic health services." *Future Generation Computer Systems* 68 (2017): 1-13.
143. Fernández-Alemán, José Luis, et al. "Security and privacy in electronic health records: A systematic literature review." *Journal of biomedical informatics* 46.3 (2013): 541-562.
144. Sajid, Anam, and Haider Abbas. "Data privacy in cloud-assisted healthcare systems: state of the art and future challenges." *Journal of medical systems* 40.6 (2016): 1-16.
145. Aggarwal, Charu C., and S. Yu Philip. *A general survey of privacy-preserving data mining models and algorithms*. Springer US, 2008.
146. Zhang, Rui, Ling Liu, and Rui Xue. "Role-based and time-bound access and management of EHR data." *Security and Communication Networks* 7.6 (2014): 994-1015.
147. Ruj, Sushmita, Milos Stojmenovic, and Amiya Nayak. "Privacy preserving access control with authentication for securing data in clouds." *Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on*. IEEE, 2012.
148. Ganz, Nicole. Data Anonymization and its Effect on Personal Privacy. Diss. State University Of New York, 2015
149. Hay, Michael, et al. "Accurate estimation of the degree distribution of private networks." 2009 Ninth IEEE International Conference on Data Mining. IEEE, 2009.
150. Karwa, Vishesh, et al. "Private analysis of graph structure." Proceedings of the VLDB Endowment 4.11 (2011): 1146-1157
151. Proserpio, Davide, Sharon Goldberg, and Frank McSherry. "A workflow for differentially-private graph synthesis." Proceedings of the 2012 ACM workshop on Workshop on online social networks. ACM, 2012
152. Andreu-Perez, Javier, et al. "Big data for health." IEEE journal of biomedical and health informatics 19.4 (2015): 1193-1208.

153. Heurix, Johannes, et al. "A taxonomy for privacy enhancing technologies." Computers & Security 53 (2015): 1-17.
154. M. Rahimi, “Extended K-Anonymity Model for Privacy Preserving on Micro Data,” *I. J. Comput. Netw. Inf. Secur.*, no. November, pp. 42– 51, 2015
155. Wagner, Isabel, and David Eckhoff. "Technical privacy metrics: a systematic survey." *ACM Computing Surveys (CSUR)* 51.3 (2018): 57.
156. X. Li, Jiexing and Tao, Yufei and Xiao, “Preservation of proximity privacy in publishing numerical sensitive data,” *Proc. 2008 ACM SIGMOD Int. Conf. Manag. data*, 2008.
157. Aurucci, Paola, et al. ""GDPR" IMPACT ON HEALTH DATA EXCHANGE IN EUROPEAN DIGITAL ENVIRONMENT." *E-HEALTH 2018 ICT, SOCIETY AND HUMAN BEINGS 2018*: 45.
158. Seol, Kwangsoo, et al. "Privacy-preserving attribute-based access control model for XML-based electronic health record system." *IEEE Access* 6 (2018): 9114-9128.